On the Out-of-Sample Performance of Stochastic Dynamic Programming and Model Predictive Control

Dominic S. T. Keehan
University of Auckland, dkee331@aucklanduni.ac.nz

Andrew B. Philpott
University of Auckland, a.philpott@auckland.ac.nz

Edward J. Anderson
Imperial College London, e.anderson@imperial.ac.uk

Sample average approximation-based stochastic dynamic programming (SDP) and model predictive control (MPC) are two different methods for approaching multistage stochastic optimization. In this paper we investigate the conditions under which SDP may be outperformed by MPC. We show that, depending on the presence of concavity or convexity, MPC can be interpreted as solving a mean-constrained distributionally-ambiguous version of the problem that is solved by SDP. This furnishes performance guarantees when the true mean is known and provides intuition for why MPC performs better in some applications and worse in others. We then study a multistage stochastic revenue optimization problem that is representative of the type for which MPC may be the better choice. We find that this may indeed be the case when the probability distribution of the underlying random variable is skewed or has enough weight in the right hand tail, and support this with a number of examples.

Key words: sample average approximation, stochastic dynamic programming, model predictive control, distributionally robust optimization

---

1. Introduction

In practice multistage stochastic optimization problems often have to be solved without explicit knowledge of the probability distributions involved. Although one can create scenario-tree approximations of such problems based on samples of the random variables in each stage (termed *sample average approximation* or SAA), the number of samples required to solve the true problem to a specified accuracy grows exponentially with the number of stages (Shapiro and Nemirovski 2005, Shapiro 2006) and solving the resulting problem is computationally expensive (Dyer and Stougie 2006). It follows that for large problems SAA may only be practically applied when the number of samples in each stage is small. We are interested in the performance of different solution approaches in this small-sample regime.

Multistage stochastic optimization becomes easier when the random variables are stage-wise independent or follow a Markov process and the problem can be formulated as a stochastic optimal control problem. In principle, such problems are amenable to solution via Stochastic Dynamic Programming (SDP), as long as the state-dimension is not too high (Bellman 1957). However, there are alternative solution approaches. Often, Model Predictive Control (MPC) is used. MPC fixes the random variables involved in the multistage stochastic optimization problem to a deterministic forecast value, giving a cost-to-go function that is computationally inexpensive to evaluate (Bertsekas 2005, 2022). By replacing each random variable with a deterministic forecast, MPC can incorporate high state dimensions, nonlinear constraints, and a large number of stages. Although the optimal solutions obtained from SDP and MPC coincide for certain quadratic problems (Theil 1957, Ziemba 1971), this is the exception rather than the rule.

In our paper both SDP and MPC are sample based. We use SAA to construct approximations of the true probability distributions, assigning equal probability to each of N random samples. Also, we use a simple form of MPC that forecasts the random variables as the average of these N samples. Compared to SDP, in certain applications MPC performs poorly out-of-sample (Pacaud et al. 2024), and in other applications MPC performs well out-of-sample (Martin 2021). This is surprising as MPC is tantamount to ignoring randomness.

The intention of this paper is to advance our understanding of SDP and MPC applied to stochastic optimal control problems, by comparing the out-of-sample performance of each method when N is small. Our study is motivated by the question:

Under what conditions is sample-based SDP outperformed by MPC out-of-sample?

To answer this question we first look through the lens of distributional ambiguity. In data-driven stochastic optimization, when N is small, considering an ambiguity set of probability distributions informed by the N samples can improve out-of-sample performance (Anderson and Philpott 2022, Gotoh et al. 2023). Moment constraints are a classical choice for constructing these ambiguity sets (Scarf 1958, Dupačová 1966, Delage and Ye 2010). Owing to Jensen's inequality, depending on the convexity or concavity of the cost-to-go function in the random variables, we show that sample-based MPC can be interpreted as solving a mean-constrained distributionally-ambiguous version of the problem that is solved by sample-based SDP.

The distributionally-ambiguous interpretations we provide are *not* guarantees that one method will outperform the other. To gain a deeper understanding of the conditions under which SDP can be outperformed by MPC out-of-sample, we study a specific multistage revenue optimization problem that determines how much product from an existing inventory to sell at a given price offered by the market. Unsold product may be stored in inventory for later sales at a different realisation of the random market price, subject to deterministic holding costs. This problem is simple enough to admit the derivation of a closed-form optimal policy, but complex enough to capture critical aspects of the differences in out-of-sample performance between the methods.

Given the revenue optimization problem and some ground-truth probability distribution for the random market price, we derive an optimal SDP policy for any N fixed samples and hence can evaluate its out-of-sample performance under the true distribution. Similarly, we derive an optimal MPC policy and can evaluate its performance under the true distribution. These two values enable us to understand the sensitivities of each method to the specific values of the N samples used in the distributional approximations. Furthermore, the expectation of these two values over the sampling distribution of the N samples quantifies the average performance of each method.

In Section 2 we formulate a general stochastic optimal control problem as well as its data-driven approximations via sample-based SDP and MPC. In Section 3 we classify the Bellman operator associated with the MPC problem as concavity or convexity preserving, and show that in the former case MPC can be interpreted as solving a mean-constrained distributionally-robust version of the problem that is solved by SDP. In the latter case this becomes a distributionally-optimistic interpretation. We then consider a number of specific examples where the MPC Bellman operator is concavity or convexity preserving. In Section 4 we show that the distributionally ambiguous interpretations developed in the previous section can be used to derive performance guarantees for MPC when the true mean is known. We also derive a result for comparing the out-of-sample performance of different policies for stochastic optimal control problems. In Section 5 we introduce our multistage revenue optimization problem and establish a closed-form expression for its optimal solution. We then compare the out-of-sample performance of SDP and MPC and provide a condition on the samples which ensures that MPC performs at least as well as SDP. In the remainder of Section 5 we report on some examples. Firstly, we show that when the underlying distribution of the market price is exponential, the expected out-of-sample performance improvement from using MPC instead of SDP becomes arbitrarily large as the discount factor approaches 1. Lastly, we present a range of numerical experiments that support the observations of the previous sections. The paper concludes with a discussion in Section 6.

All of the proofs of the results in this paper are deferred to the appendices.

2. Stochastic optimal control

Before introducing the control problem, we introduce some notation. An element $x \in \mathcal{X} \subset \mathbb{R}^n$ denotes the "state" vector in the multistage optimization problem. An element $\xi \in \Xi \subset \mathbb{R}^m$ denotes the value of a random "noise" vector $\tilde{\xi}$ distributed according to a probability distribution P. We use tildes to denote random vectors. All of the probability distributions considered in this paper are Borel measurable functions from the Borel σ-algebra on Ξ (or those of its products) to that on $\mathbb{R}$. We use superscripts to denote indexing by stage, but also, where clear from context, to denote raising to powers.

For an initial state $x^1 \in \mathcal{X}$, we study the infinite-horizon stochastic optimal control problem

$$\begin{aligned} \underset{y^1,y^2,\ldots}{\text{minimize}} \quad & \mathbb{E}_{P^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \tilde{\xi}^t) \in \mathcal{Y}(x^t, \tilde{\xi}^t) \quad \forall t \in \mathbb{N} \end{aligned} \tag{SOC}$$

in which $\{y^t : \mathcal{X} \times \Xi \mapsto \mathcal{X},\, t \in \mathbb{N}\}$ is a *policy* that at each stage t provides a *decision rule* y^t for the next state vector x^{t+1} given the current state x^t and the realization of the random vector $\tilde{\xi}^t$. At each stage the map $\xi \mapsto y^t(x, \xi)$ is required to be measurable for each $x \in \mathcal{X}$. Here $\beta \in (0, 1)$ is a discount factor (with β^t its t^{th} power), $\varphi : \mathcal{X} \times \mathcal{X} \times \Xi \mapsto \mathbb{R}$ is a cost function, and $\mathcal{Y} : \mathcal{X} \times \Xi \mapsto 2^{\mathcal{X}}$ is a set-valued function enforcing state-transition constraints. The infinite-product probability distribution $P^\infty := P \times P \times \cdots$ defines a joint distribution for the random vectors $\tilde{\xi}^1, \tilde{\xi}^2, \ldots$.

In this formulation we have assumed that the dynamics of the state evolution is such that the control is exercised after the noise is known. For that reason we can take the control as simply the choice of state at the next period. The noise has an effect on the possible choice of x^{t+1} as well as on the cost function.

In general, (SOC) may not be well-defined. To keep our analysis simple, we make the following assumption.

Assumption 1.

(i) *The set $\mathcal{X}$ is closed and convex, and the set-valued function $\mathcal{Y} : \mathcal{X} \times \Xi \mapsto 2^{\mathcal{X}}$ is nonempty- and compact-valued, and continuous.*

(ii) *The set Ξ is closed and convex, and the random vectors $\tilde{\xi}^1, \tilde{\xi}^2, \ldots$ are independent and identically distributed according to P.*

(iii) *The function $\varphi : \mathcal{X} \times \mathcal{X} \times \Xi \mapsto \mathbb{R}$ is bounded and continuous.*

Note that Assumption 1 (ii) does not preclude modelling stagewise-dependent randomness, as auxiliary states that track the realizations of random vectors in previous stages can be created, as we do in Example 1 below. Moreover, the boundedness condition in Assumption 1 (iii) can often be relaxed, although the mathematical tools required become more complex.

For simplicity we do not vary P, φ, or $\mathcal{Y}$ between stages. However, it will be clear how our results can be generalized to such a setting using a periodic formulation of (SOC), as in Shapiro and Ding 2020. Furthermore, in this paper we consider only infinite-horizon stochastic optimal control problems, although all of our results have finite-horizon counterparts. For a comparison of the finite- and infinite-horizon cases, see Bertsekas and Shreve 1996, Chapters 3 and 4.

The problem (SOC) is an infinite-dimensional optimization problem. A closely related problem is that of finding a function $v : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ which solves the functional equation

$$v(x,\xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \left\{ \varphi(x,y,\xi) + \beta \mathbb{E}_P \left[v(y,\tilde{\xi}) \right] \right\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{1}$$

Under Assumption 1 the equation (1) has a unique bounded and continuous solution $v : \mathcal{X} \times \Xi \mapsto \mathbb{R}$. Moreover, the value $\mathbb{E}_P \left[v(x^1,\tilde{\xi}) \right]$ is equal to the optimal value of (SOC), and there is a function $y : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ with $\xi \mapsto y(x,\xi)$ measurable for each $x \in \mathcal{X}$ satisfying

$$y(x,\xi) \in \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \left\{ \varphi(x,y,\xi) + \beta \mathbb{E}_P \left[v(y,\tilde{\xi}) \right] \right\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi.$$

This defines a decision rule for each stage t and the resulting policy solves (SOC). For a discussion of all of these well-known results, see e.g. Stokey et al. 1989, Chapter 9 or Bertsekas and Shreve 1996, Chapter 4. We refer to any function $y : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ with $\xi \mapsto y(x,\xi)$ measurable for each $x \in \mathcal{X}$ and $y(x,\xi) \in \mathcal{Y}(x,\xi)$ for each $(x,\xi) \in \mathcal{X} \times \Xi$ as a *feasible* policy for (SOC).

2.1. Data-driven stochastic optimal control

To each of N samples $\xi_1, \ldots, \xi_N \in \Xi$ from the true distribution P we allocate probability $1/N$ to construct the empirical probability distribution P_N. The infinite-product probability distribution $P_N^\infty := P_N \times P_N \times \cdots$ defines a joint distribution for the random vectors $\tilde{\xi}^1, \tilde{\xi}^2, \ldots$ which leads to a sample-based approximation of (SOC):

$$\begin{aligned} \operatorname*{minimize}_{y^1, y^2, \ldots} \quad & \mathbb{E}_{P_N^\infty} \left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x^t, x^{t+1}, \tilde{\xi}^t) \right] \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \tilde{\xi}^t) \in \mathcal{Y}(x^t, \tilde{\xi}^t) \quad \forall t \in \mathbb{N}. \end{aligned} \tag{SSOC}$$

To solve (SSOC), P is replaced by P_N in (1), yielding the problem of finding a function $v_{\mathrm{S}} : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ which solves the sample-based SDP functional equation

$$v_{\mathrm{S}}(x,\xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \left\{ \varphi(x,y,\xi) + \beta \mathbb{E}_{P_N} \left[v_{\mathrm{S}}(y,\tilde{\xi}) \right] \right\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{2}$$

Under Assumption 1 the same remarks regarding uniqueness, boundedness, continuity, and measurability made with reference to (1) apply to (2), and by solving (2) a sample-based SDP policy $y_{\mathrm{S}} : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ satisfying

$$y_{\mathrm{S}}(x,\xi) \in \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \left\{ \varphi(x,y,\xi) + \beta \mathbb{E}_{P_N} \left[v_{\mathrm{S}}(y,\tilde{\xi}) \right] \right\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi$$

can be obtained.

MPC proceeds in a different way; in (1) the random vector $\tilde{\xi}$ is replaced by a deterministic vector which is our forecast for $\tilde{\xi}$. Given a sample $\xi_1, \ldots, \xi_N$, we use the sample average $\mu_N := \frac{1}{N}\sum_{i=1}^N \xi_i \in \Xi$ as the forecast. This yields the problem of finding a function $v_{\mathrm{M}} : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ which solves the sample-based MPC functional equation

$$v_{\mathrm{M}}(x,\xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta v_{\mathrm{M}}(y,\mu_N) \Big\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{3}$$

Observe that (3) is equivalent to (1) with the probability measure δ_{μ_N} used instead of P. (Here we use δ_ξ to denote the point-mass probability distribution that assigns probability 1 to $\xi \in \Xi$.) Under Assumption 1, the same remarks regarding uniqueness, boundedness, continuity, and measurability made with reference to (1) apply to (3), and by solving (3) a sample-based MPC policy $y_{\mathrm{M}} : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ satisfying

$$y_{\mathrm{M}}(x,\xi) \in \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta v_{\mathrm{M}}(y,\mu_N) \Big\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi$$

can be obtained. We are interested in comparing the performance of the sample-based SDP and MPC policies when applied out-of-sample to the true problem (SOC).

3. Connections with distributionally ambiguous optimization

The SDP policy obtained by solving (2) depends only on the N-sample empirical probability distribution. Instead, by hedging against the worst-case probability distributions in an ambiguity set $\mathcal{P}$ informed by the N samples, *distributionally robust optimization* (DRO) provides a way to limit out-of-sample disappointment. The infinite product of these ambiguity sets $\mathcal{P}^\infty := \mathcal{P} \times \mathcal{P} \times \cdots$ defines a set of joint distributions for the random vectors $\tilde{\xi}^1, \tilde{\xi}^2, \ldots$, which leads to a distributionally robust version of (SSOC);

$$\begin{aligned} \underset{y^1, y^2, \ldots}{\text{minimize}} \quad & \sup_{R \in \mathcal{P}^\infty} \mathbb{E}_R \left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x^t, x^{t+1}, \tilde{\xi}^t) \right] \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \tilde{\xi}^t) \in \mathcal{Y}(x^t, \tilde{\xi}^t) \quad \forall t \in \mathbb{N}. \end{aligned} \tag{DR-SSOC}$$

The choice of $\mathcal{P}^\infty$ in (DR-SSOC) requires care. By using a product set, under the assumption of stage-wise independence, $\mathcal{P}^\infty$ satisfies the “rectangularity” property (see Iyengar 2005, Assumption 2.1 and Shapiro 2016 for details), whereby (DR-SSOC) can be solved by finding a function $v_{\mathrm{R}} : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ which satisfies the DRO functional equation

$$v_{\mathrm{R}}(x,\xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \sup_{Q \in \mathcal{P}} \mathbb{E}_Q \big[v_{\mathrm{R}}(y,\tilde{\xi}) \big] \Big\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{4}$$

Note that the ambiguity set present in (4) is $\mathcal{P}$ so the supremum (4) is over a set of distributions for $\tilde{\xi}$, rather than a set of joint distributions for $\tilde{\xi}^1, \tilde{\xi}^2, \ldots$ as in (DR-SSOC).

To demonstrate a connection between DRO and MPC requires analysis of the functional operator associated with (3). Let $\mathcal{C}(\mathcal{X}\times\Xi)$ denote the set of bounded and continuous $\mathbb{R}$-valued functions on $\mathcal{X}\times\Xi$. This is a Banach space under the sup norm $\|\cdot\|_\infty$. Under Assumption 1, the equation (3) features an MPC Bellman operator $B_{\mathrm{M}}:\mathcal{C}(\mathcal{X}\times\Xi)\mapsto\mathcal{C}(\mathcal{X}\times\Xi)$ which for $f\in\mathcal{C}(\mathcal{X}\times\Xi)$ at $(x,\xi)\in\mathcal{X}\times\Xi$ has value

$$B_{\mathrm{M}}(f)(x,\xi) \coloneqq \inf_{y\in\mathcal{Y}(x,\xi)}\big\{\varphi(x,y,\xi)+\beta f(y,\mu_N)\big\}.$$

The operator B_{M} can be defined for any forecast in place of the sample average μ_N which appears here. For a proof of the fact that B_{M} maps $\mathcal{C}(\mathcal{X}\times\Xi)$ into itself, see e.g. Stokey et al. 1989, Theorem 9.6.

Recall that B_{M} is a contraction mapping under the sup norm, since

$$\|B_{\mathrm{M}}(f)-B_{\mathrm{M}}(g)\|_\infty \le \beta\|f-g\|_\infty \quad \forall f,g\in\mathcal{C}(\mathcal{X}\times\Xi)$$

with $\beta\in(0,1)$. Banach's Fixed-Point Theorem then shows that the equation $f=B_{\mathrm{M}}(f)$ has a unique solution in $(\mathcal{C}(\mathcal{X}\times\Xi),\|\cdot\|_\infty)$, and this solution solves the MPC functional equation (3). Furthermore, if there exists a closed subset $\mathcal{F}\subset\mathcal{C}(\mathcal{X}\times\Xi)$ which maps to itself under B_{M} and $f\in\mathcal{C}(\mathcal{X}\times\Xi)$ solves $f=B_{\mathrm{M}}(f)$, then it holds that $f\in\mathcal{F}$ (Stokey et al. 1989, Theorem 3.2, Corollary 1). This motivates the following definition.

Definition 1. The MPC Bellman operator B_{M} is *concavity preserving* (respectively *convexity preserving*) if there exists a closed subset $\mathcal{F}\subset\mathcal{C}(\mathcal{X}\times\Xi)$ where the map $\xi\mapsto f(x,\xi)$ is concave (convex) for each $x\in\mathcal{X}$ and $f\in\mathcal{F}$, and which maps to itself under B_{M}.

When the state-transition constraints do not depend on the random terms and the cost function is concave in the random terms, the MPC Bellman operator is concavity preserving.

Proposition 1. *Let Assumption 1 hold. Suppose that the set-valued function $\mathcal{Y}(x)$ does not depend on ξ and that the map $\xi\mapsto\varphi(x,y,\xi)$ is concave for each $x,y\in\mathcal{X}$. Then the MPC Bellman operator B_{M} is concavity preserving.*

When the MPC Bellman operator is concavity preserving, we have the following DRO interpretation.

Theorem 1. *Let Assumption 1 hold. For an ambiguity set $\mathcal{P}$, suppose that $\mathbb{E}_Q[\tilde{\xi}]=\mu_N$ for each $Q\in\mathcal{P}$ and that $\delta_{\mu_N}\in\mathcal{P}$. If the MPC Bellman operator B_{M} is concavity preserving, then the DRO functional equation (4) has a solution which is the same as the solution to the MPC functional equation (3).*

In contrast to DRO, *distributionally optimistic optimization* (DOO) considers only the best-case probability distributions in the ambiguity set $\mathcal{P}$. Using the rectangular set $\mathcal{P}^\infty$, a distributionally optimistic version of (SSOC) may be formed. As for (DR-SSOC), this problem can be solved by finding a function $v_{\mathrm{O}} : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ which satisfies the DOO functional equation

$$v_{\mathrm{O}}(x,\xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \inf_{Q \in \mathcal{P}} \mathbb{E}_Q \big[v_{\mathrm{O}}(y,\tilde{\xi}) \big] \Big\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{5}$$

The essential characteristic of the connection between DRO and MPC is the concavity preserving feature of the operator B_{M}. If instead B_{M} is convexity preserving, the following converse result to Theorem 1 holds.

Corollary 1. *Suppose that the conditions in the statement of Theorem 1 hold, but that the MPC Bellman operator B_{M} is convexity preserving. Then the DOO functional equation (5) has a solution which is the same as the solution to the MPC functional equation (3).*

We now give two examples with concavity or convexity preserving MPC Bellman operators, that do not satisfy the conditions of Proposition 1.

Example 1. Downward et al. (2020) study a class of problems where stagewise-dependent randomness appearing in the cost function is modelled using an autoregressive process. Here we consider a specific problem from this class. We consider a problem in which stage costs in a state y are given by $C(y) - e^{\eta}(x-y)$ where x is the previous state, and η is a random variable which follows a lag-1 autoregressive process with parameter α. We can deal with this through an additional state variable $\rho \in \mathcal{R}$. This gives the functional equation

$$v(x,\rho,\xi) = \inf_{(y,\eta) \in \mathcal{Y}(x,\rho,\xi)} \Big\{ C(y) - e^{\eta}(x-y) + \beta \mathbb{E}\big[v(y,\eta,\tilde{\xi}) \big] \Big\} \quad \forall (x,\varrho,\xi) \in \mathcal{X} \times \mathcal{R} \times \Xi$$

where

$$\mathcal{Y}(x,\rho,\xi) := \big\{ (y,\eta) \in \mathcal{X} \times \mathcal{R} \,\big|\, 0 \le y \le x,\ \eta = \alpha\rho + \xi \big\}.$$

To see that the resulting MPC Bellman operator is concavity preserving, consider the set

$$\mathcal{F} := \big\{ f \in \mathcal{C}(\mathcal{X} \times \mathcal{R} \times \Xi) \,\big|\, (\rho,\xi) \mapsto f(x,\rho,\xi) \text{ is concave for each } x \in \mathcal{X} \big\}.$$

Then we can check that for each $f \in \mathcal{F}$ and $y \le x$ both the terms $-e^{\eta}(x-y)$ and $\mathbb{E}\big[f(y,\eta,\tilde{\xi})\big]$ are concave in ϱ and ξ. Since the infimum operator preserves concavity we have shown that, for this problem, sample-based MPC satisfies the distributionally-robust interpretation of Theorem 1.

Example 2. The *hydrothermal scheduling* problem (Pereira and Pinto 1991) minimizes the thermal energy cost of meeting energy demand in a system with hydroelectric plants supplied by reservoir storage with random inflows. Consider a thermal plant with generation t and marginal cost c, one hydro plant that generates $d-t$ where d is a constant demand per period, and a reservoir with energy storage x and random inflow $\tilde{\xi}$. All quantities are measured in terms of energy. This gives the following functional equation for the infinite horizon discounted cost of meeting demand in each period

$$v(x,\xi) = \inf_{(y,t)\in\mathcal{Y}(x,\xi)} \left\{ ct + \beta \mathbb{E}\big[v(y,\tilde{\xi})\big] \right\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi,$$

where

$$\mathcal{Y}(x,\xi) := \big\{ (y,t) \in \mathcal{X} \times \mathcal{T} \,\big|\, y = x + t - d + \xi \big\}.$$

To see that the resulting MPC Bellman operator is convexity preserving, consider the set

$$\mathcal{F} := \big\{ f \in \mathcal{C}(\mathcal{X} \times \Xi) \,\big|\, (x,\xi) \mapsto f(x,\xi) \text{ is convex} \big\}.$$

where $\mathcal{T}$ is a compact convex set. The proof is similar to that of the proof of Proposition 1. Hence, for this problem, sample-based MPC satisfies the distributionally-optimistic interpretation of Corollary 1.

If the MPC Bellman operator is concavity preserving, Theorem 1 shows that the sample-based MPC policy obtained by solving the MPC functional equation (3) may be interpreted as distributionally robust. Distributional robustness can improve out-of-sample performance due to the effects of shrinkage (Anderson and Philpott 2022). Conversely, if the MPC Bellman operator is convexity preserving, Corollary 1 shows that the sample-based MPC policy may be interpreted as distributionally optimistic. Distributional optimism can worsen out-of-sample performance due to increased decision variance (Gotoh et al. 2023). This suggests that for small N, when the MPC Bellman operator is concavity preserving (as in Proposition 1 and Example 1), MPC may be a better choice than SDP. Conversely, for small N, when the MPC Bellman operator is convexity preserving (as in Example 2), SDP may be a better choice than MPC.

4. Performance guarantees and out-of-sample comparisons

Suppose that the probability distribution P has a finite mean $\mathbb{E}_P[\tilde{\xi}]$, and let $\tilde{\xi}_1,\ldots,\tilde{\xi}_N$ be N random samples which are independent and identically distributed according to P. Generally, the sample average $\frac{1}{N}\sum_{i=1}^N \tilde{\xi}_i$ is a good estimate for $\mathbb{E}_P[\tilde{\xi}]$, as long as N is sufficiently large. In fact, the Strong Law of Large Numbers shows that $\frac{1}{N}\sum_{i=1}^N \tilde{\xi}_i \to \mathbb{E}_P[\tilde{\xi}]$ as $N \to \infty$ almost surely. MPC can take advantage of this situation since the computational complexity of solving (3) does not depend on the size of N. SDP of course, cannot.

We now show that, when the true mean is known and used as the MPC forecast, the solution to the MPC functional equation provides a performance guarantee on the cost incurred when applying the resulting MPC policy to (SOC); that is, we will do better than predicted.

Theorem 2. *Let Assumption 1 hold, and set $\mu_N = \mu := \mathbb{E}_P[\tilde{\xi}]$ in the MPC functional equation (3). If the resulting MPC Bellman operator B_M is concavity preserving, then*

$$\mathbb{E}\left[\sum_{t=1}^{\infty}\beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \le v_{\mathrm{M}}(x^1, \mu) \quad \textit{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N};$$

that is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (3) is an upper-bound on the cost incurred when applying the MPC policy to (SOC).

If N is large then using the sample-based form of MPC with μ_N will give a result that is close to that with the true mean μ, which suggests that applying the sample-based MPC policy to (SOC) is unlikely to result in out-of-sample disappointment when the MPC Bellman operator is concavity preserving. We also have a lower bound on costs in the case where the MPC Bellman operator is convexity preserving.

Corollary 2. *Suppose that the conditions in the statement of Theorem 2 hold, but that the MPC Bellman operator is convexity preserving. Then*

$$\mathbb{E}\left[\sum_{t=1}^{\infty}\beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \ge v_{\mathrm{M}}(x^1, \mu) \quad \textit{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N};$$

that is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (3) is a lower-bound on the cost incurred when applying the MPC policy to (SOC).

In the rest of this section we derive a result for comparing the cost incurred when applying different policies to (SOC). Given a policy $y : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ that is feasible for (SOC), with the function φ continuous the map $\xi \mapsto \varphi(x, y(x, \xi), \xi)$ is measurable for each $x \in \mathcal{X}$. Hence, the objective value of (SOC) evaluated under this policy is well defined. Let $\boldsymbol{V}_y : \mathcal{X} \mapsto \mathbb{R}$ denote this value as a function of the initial state. For each $x^1 \in \mathcal{X}$ we have that

$$\boldsymbol{V}_y(x^1) := \mathbb{E}\left[\sum_{t=1}^{\infty}\beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \quad \text{where } x^{t+1} = y(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N}. \tag{6}$$

If y is constructed from N random samples, then $\boldsymbol{V}_y(x^1)$ is the *out-of-sample* cost of y.

When studying out-of-sample performance, we do not require the boundedness condition of Assumption 1 (iii). Instead, we make the integrability assumption

Assumption 2. *Assumption 1 (i) and (ii) hold, and*

(iii) *The function $\varphi : \mathcal{X} \times \mathcal{X} \times \Xi \mapsto \mathbb{R}$ is continuous, and there exists a positive-valued random variable $L(\tilde{\xi})$ with $\mathbb{E}_P\big[L(\tilde{\xi})\big] < \infty$, such that for each $x, y \in \mathcal{X}$ and P-almost every $\xi \in \Xi$ it holds that $L(\xi) \ge |\varphi(x, y, \xi)|$.*

Under Assumption 2 (iii), each term in the sum in (6) is bounded by the random variable $L(\tilde{\xi}^t)$ which has a finite expectation. With $\beta \in (0,1)$, it follows that $\boldsymbol{V}_y$ is bounded. Having defined $\boldsymbol{V}_y$ as a function of the initial state and shown that it is bounded, it satisfies the functional equation

$$\boldsymbol{V}_y(x) = \mathbb{E}_P\left[\varphi\big(x, y(x,\tilde{\xi}), \tilde{\xi}\big) + \beta \boldsymbol{V}_y\big(y(x,\tilde{\xi})\big)\right]. \tag{7}$$

Our approach to compare two different policies is to consider starting with one policy and then switching to the other policy after a certain number of stages. To this end, we make the following definition.

Definition 2. Given two policies $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ which are feasible for (SOC), define the function $\boldsymbol{V}_{yy'} : \mathcal{X} \mapsto \mathbb{R}$ by

$$\boldsymbol{V}_{yy'}(x) := \mathbb{E}_P\left[\varphi\big(x, y(x,\tilde{\xi}), \tilde{\xi}\big) + \beta \boldsymbol{V}_{y'}\big(y(x,\tilde{\xi})\big)\right].$$

The value $\boldsymbol{V}_{yy'}(x)$ is the objective value of (SOC) starting from initial state $x \in \mathcal{X}$ if the policy y is used for the first stage and then the policy y' is used thereafter. It is clear that $\boldsymbol{V}_{yy'}$ is well-defined and bounded in the same way that $\boldsymbol{V}_y$ and $\boldsymbol{V}_{y'}$ are.

Theorem 3. *Let Assumption 2 hold. Suppose that $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ are two feasible policies for (SOC). If*

$$\boldsymbol{V}_{y'}(x) \le \boldsymbol{V}_{yy'}(x) \quad \forall x \in \mathcal{X},$$

then $\boldsymbol{V}_{y'}(x) \le \boldsymbol{V}_y(x)$ for each $x \in \mathcal{X}$.

Rather than having to calculate integrals directly to compare the out-of-sample performance of different policies, Theorem 3 allows us to check a uniform condition involving functional equations with similar definitions.

5. Revenue optimization with stochastic prices

To study the out-of-sample performance of sample-based SDP and MPC, we will look at a particular problem within the stochastic optimal control setting developed in Section 2. Let $\mathbb{R}_+$ denote the nonnegative real numbers. For an initial inventory level $x^1 \in \mathbb{R}_+$, we formulate the revenue optimization problem

$$\begin{aligned} \underset{y^1, y^2, \dots}{\text{maximize}} \quad & \mathbb{E}_{P^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1}\big(\tilde{p}^t(x^t - x^{t+1}) - C(x^{t+1})\big)\right] \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \tilde{p}^t) \in [0, x^t] \quad \forall t \in \mathbb{N}. \end{aligned} \tag{ROP}$$

As with the general problem (SOC), a decision at stage t is made for the next inventory level x^{t+1} given the current inventory level x^t and the realization of $\tilde{p}^t$. Here $\tilde{p}^1, \tilde{p}^2, \dots$ are random per-unit sales prices (henceforth prices) that are independent and identically distributed according to P, and $C:\mathbb{R}_+ \mapsto \mathbb{R}_+$ is an inventory storage cost function. We require C to be increasing, strictly convex, and continuously differentiable. Writing $c(x) \coloneqq \frac{\mathrm{d}}{\mathrm{d}x}C(x)$, because c is a strictly increasing continuous function, we may also define an inverse function c^{-1} on the range of c which is continuous as well. Since the inventory levels for (ROP) are restricted to the compact set $[0, x^1]$, without loss of generality we assume that C is such that $\lim_{x\to\infty} c(\infty) = \infty$.

The problem (ROP) can be interpreted as that facing a merchant who maximizes their expected discounted reward by at each time t selling down from their current inventory level x^t to a new inventory level x^{t+1} at a realization of the random price $\tilde{p}^t$, while incurring a storage cost $C(x^{t+1})$ on their remaining inventory. This model may be applied in a number of situations: for example the model could describe an electricity distributor with a charged battery needing to decide when to sell electricity, or an investor who needs to decide when to sell a holding of shares.

As in (SOC), the problem (ROP) is closely related to that of finding a function $v:\mathbb{R}_+\times\mathbb{R}\mapsto\mathbb{R}$ which solves the functional equation

$$v(x,p) = \sup_{0\le y\le x}\left\{p(x-y) - C(y) + \beta\mathbb{E}_P\big[v(y,\tilde{p})\big]\right\} \quad \forall (x,p)\in\mathbb{R}_+\times\mathbb{R}. \tag{8}$$

Observing that the constraints in (ROP) restrict states to within the compact set $[0,x^1]$, when P has compact support the terms in the sum in (ROP) are bounded. Using the results of (Stokey et al. 1989, Chapter 9) the equation (8) has a unique continuous solution $v:\mathbb{R}_+\times\mathbb{R}\mapsto\mathbb{R}$ and the map $x\mapsto v(x,p)$ is concave for each $p\in\mathbb{R}$. Moreover, given a solution v, a policy $y:\mathbb{R}_+\times\mathbb{R}\mapsto\mathbb{R}_+$ satisfying

$$y(x,p) \in \operatorname*{argmax}_{0\le y\le x}\left\{p(x-y) - C(y) + \beta\mathbb{E}_P\big[v(y,\tilde{p})\big]\right\} \quad \forall (x,p)\in\mathbb{R}_+\times\mathbb{R} \tag{9}$$

is an optimal policy for (ROP).

A suitable change of perspective from maximization to minimization in ROP yields

$$\varphi(x,y,p) = -p(x-y) + C(y)$$

which is concave in p. Therefore ROP satisfies the conditions of Proposition 1, and applying sample-based MPC to (ROP) will yield the distributionally robust outcome provided by Theorem 1.

Denote the projection of $z\in\mathbb{R}$ onto the closed interval $[a,b]\subset\mathbb{R}$ by $(z)_{[a,b]} \coloneqq \max\{a,\min\{z,b\}\}$, and let $(z)_+ \coloneqq \max\{0,z\}$. We have the following closed-form expression for an optimal policy for (ROP).

Proposition 2. *Suppose that the probability distribution P has compact support. Then the policy $y:\mathbb{R}_+\times\mathbb{R}\mapsto\mathbb{R}_+$ which for inventory x and price p sells down to the inventory level*

$$y(x,p) = c^{-1}\left(\Big(\beta\mathbb{E}_P\big[(\tilde{p}-p)_+\big] - (1-\beta)p\Big)_{[c(0),c(x)]}\right)$$

is optimal for (ROP).

The optimal policy has a natural interpretation: the term $\beta \mathbb{E}_P\left[(\tilde{p}-p)_+\right]$ is the discounted expected increase in price gained by not selling, and the term $(1-\beta)p$ is the portion of the current price that is irrecoverably lost due to discounting by not selling. The difference in these terms gives the expected net-increase in price gained from storing inventory, and this is balanced against the marginal cost of storage.

Proposition 2 shows that for each price p, the problem (ROP) has an optimal target inventory level $y(\infty,p) := \lim_{x\to\infty} y(x,p)$. At inventory level x, the optimal policy is to sell down to $y(\infty,p)$ if x is above this, and sell nothing otherwise. Conversely, there is a minimal price $p^*(x)$ required for selling any portion of inventory from x to be worthwhile: this is the highest price p which solves $y(\infty,p)=x$. Note that due to the continuity of the function c^{-1}, the map $x \mapsto p^*(x)$ is continuous.

Proposition 2 makes no assumptions about the probability distribution P, except that it has compact support. Thus, P could have a density on a compact set, or P could be an empirical probability distribution. If N price samples $p_1,\dots,p_N$ are each assigned probability $1/N$, Proposition 2 shows that the optimal SDP policy, which we denote by y_{S}, is given by

$$y_{\mathrm{S}}(x,p) = c^{-1}\left(\left(\beta \tfrac{1}{N}\sum_{i=1}^N (p_i-p)_+ - (1-\beta)p\right)_{[c(0),c(x)]}\right). \tag{10}$$

The optimal MPC policy, which we denote by y_{M}, can then be obtained from Proposition 2 by applying it to the point-mass probability distribution at the sample average $\mu_N := \sum_{i=1}^N p_i$, giving

$$y_{\mathrm{M}}(x,p) = c^{-1}\left(\left(\beta(\mu_N-p)_+ - (1-\beta)p\right)_{[c(0),c(x)]}\right). \tag{11}$$

For the SDP and MPC policies defined by (10) and (11) respectively, we denote by $y_{\mathrm{S}}(\infty,p)$ and $y_{\mathrm{M}}(\infty,p)$ their target inventory levels at price p, and by $p^*_{\mathrm{S}}(x)$ and $p^*_{\mathrm{M}}(x)$ their minimal prices required for sales from inventory x. Clearly it holds that $p^*_{\mathrm{S}}(x) \le \max\{p_1,\dots,p_N\}$, since this is the highest price the SDP policy considers a possibility. Similar remarks hold for the MPC policy.

Depending on the price samples $p_1,\dots,p_N$, the sample-based policies may hold on to inventory for too long and overpay in storage costs, or sell inventory prematurely and not be able to take advantage of higher prices offered in the future. Jensen's inequality shows $\mathbb{E}\left[(\tilde{p}-p)_+\right] \ge (\mathbb{E}[\tilde{p}]-p)_+$, and hence $y_{\mathrm{S}}(\infty,p) \ge y_{\mathrm{M}}(\infty,p)$. Thus, the MPC policy decides that it is worthwhile to sell at lower prices than the SDP policy does. It is clear then that $p^*_{\mathrm{S}}(x) \ge p^*_{\mathrm{M}}(x)$, and the trade-off between overpaying for storage or selling prematurely is handled differently by SDP and MPC.

5.1. Out-of-sample performance

To study the out-of-sample performance of the sample-based SDP and MPC policies obtained from Proposition 2, we make the following assumption regarding the true probability distribution.

Assumption 3. *The probability distribution P has nonnegative support, no atoms, and a finite mean.*

Under Assumption 3, for each $x, y \in [0, x^1]$ and P-almost every p it holds that

$$|p(x-y) - C(y)| \leq px^1 + C(x^1),$$

with $\mathbb{E}_P\big[\tilde{p}x^1 + C(x^1)\big]$ finite valued. This shows that Assumption 3 for (ROP) is equivalent to Assumption 2 for (SOC), which enables us to define the out-of-sample performance of policies for (ROP) using functional equations.

Given N price samples $p_1, \ldots, p_N$ we form the sample-based SDP and MPC policies determined by (10) and (11) respectively. Under Assumption 3, the out-of-sample performance of SDP when starting from initial inventory level x (which we denote by $\boldsymbol{V}_{\mathrm{S}}(x)$) is well defined and satisfies the functional equation

$$\boldsymbol{V}_{\mathrm{S}}(x) = \mathbb{E}_P\Big[\tilde{p}\big(x - y_{\mathrm{S}}(x,\tilde{p})\big) - C\big(y_{\mathrm{S}}(x,\tilde{p})\big) + \beta \boldsymbol{V}_{\mathrm{S}}\big(y_{\mathrm{S}}(x,\tilde{p})\big)\Big]. \tag{12}$$

Similarly, the out-of-sample performance of MPC when starting from the initial inventory level x (which we denote by $\boldsymbol{V}_{\mathrm{M}}(x)$) is well defined and satisfies the functional equation

$$\boldsymbol{V}_{\mathrm{M}}(x) = \mathbb{E}_P\Big[\tilde{p}\big(x - y_{\mathrm{M}}(x,\tilde{p})\big) - C\big(y_{\mathrm{M}}(x,\tilde{p})\big) + \beta \boldsymbol{V}_{\mathrm{M}}\big(y_{\mathrm{M}}(x,\tilde{p})\big)\Big]. \tag{13}$$

We use Theorem 3 to establish the following proposition which provides conditions that guarantee $\boldsymbol{V}_{\mathrm{M}}(x) \geq \boldsymbol{V}_{\mathrm{S}}(x)$.

Proposition 3. *Let Assumption 3 hold and suppose that P has a probability density function $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$. For N price samples $p_1, \ldots, p_N$ which determine the sample-based SDP and MPC policies by (10) and (11), respectively, if*

$$c(x) \geq \beta \int_{p^*_{\mathrm{S}}(x)}^{\infty} pf(p)\mathrm{d}p \quad \forall x \in [0, x^1],$$

then $\boldsymbol{V}_{\mathrm{M}}(x) \geq \boldsymbol{V}_{\mathrm{S}}(x)$ for each $x \in [0, x^1]$. That is, the MPC policy performs at least as well out-of-sample on (ROP) as the SDP policy.

Proposition 3 requires that $c(x) \geq \beta \int_{p^*_{\mathrm{S}}(x)}^{\infty} pf(p)\mathrm{d}p$ for each $x \in [0, x^1]$. Without loss of generality, assume $p_1 \leq \cdots \leq p_N$ and recall $p^*_{\mathrm{S}}(x) \leq p_N$. Hence, the map $p_N \mapsto (p_N - p)_+$ in (10) is strictly increasing at the point $p = p^*_{\mathrm{S}}(x)$, where it holds that $y_{\mathrm{S}}(\infty, p) = x$. It follows that $p_N \mapsto p^*_{\mathrm{S}}(x)$ is strictly increasing, and $p_N \mapsto \int_{p^*_{\mathrm{S}}(x)}^{\infty} pf(p)\mathrm{d}p$ is then decreasing. If the N price samples $\tilde{p}_1, \ldots, \tilde{p}_N$ are independent and identically distributed according to P, when f has a small amount of probability density at very high prices we will occasionally sample a $\tilde{p}_N$ that is sufficiently large for the inequality $c(x) \geq \beta \int_{p^*_{\mathrm{S}}(x)}^{\infty} pf(p)\mathrm{d}p$ to hold for each $x \in [0, x^1]$.

Proposition 3 is not an explicit statement about the relative *expected* out-of-sample performance of sample-based SDP and MPC under the sampling distribution. In fact, for most applications it is likely that there will always be some possible realisations of the random samples which result in MPC outperforming SDP, so Proposition 3 is not surprising. However, the result demonstrates how the relative performance of SDP and MPC is affected by the parameterization of (ROP): the SDP policy can be misled by overly-high price samples which cause it to hold on to inventory for too long and overpay in storage costs. Consequently, right skew and the size of the tail in the true distribution are likely to negatively impact the performance of SDP. On the other hand, due to its distributionally robust properties, the MPC policy more quickly sells down to lower inventory levels and is protected against this.

5.2. Expected out-of-sample performance

In the rest of this section we present a number of examples which compare the expected out-of-sample performance of sample-based SDP and MPC on (ROP) for price distributions with different skews and tail sizes. We suppose that the N price samples $\tilde{p}_1, \ldots, \tilde{p}_N$ used to determine the sample-based policies by (10) and (11) are independent and identically distributed according to P. Expected out-of-sample performance is then given by the expectation of (12) and (13) under the sampling distribution of $\tilde{p}_1, \ldots, \tilde{p}_N$.

Let Exponential(λ) be an exponential distribution with rate λ. Exponential distributions are strongly right-skewed. When prices are exponentially distributed, we have the following result.

Proposition 4. *Let P be an* Exponential(1) *distribution, $x^1 := 1$, and $C(x) := \frac{1}{2}x^2$. For each sample size $N \geq 2$, as the discount factor $\beta \to 1$ the expected out-of-sample performance of sample-based SDP on (ROP) is unbounded below while the expected out-of-sample performance of sample-based MPC on (ROP) is bounded.*

Proposition 4 shows that for each sample size $N \geq 2$ the expected out-of-sample performance of sample-based SDP can be made arbitrarily worse than that of sample-based MPC by choosing a discount factor β that is sufficiently close to 1. For a particular β the performance of SDP may be improved by increasing N, but N may be required to be very large for SDP to outperform MPC. The result shows that the performance of SDP can be arbitrarily bad even for arbitrarily large N. Moreover, in real applications the sample-based SDP policy must be computed numerically, but algorithms for doing this cannot handle large N. MPC does not have this problem.

For a given sample size N_0, Proposition 4 also shows that we can make the expected out-of-sample profit of SDP negative by choosing a discount factor β sufficiently close to 1. Now consider the relationship of expected profit to sample size, N, for this value of β. The expected out-of-sample profit matches that of MPC and turns out to be positive when $N = 1$; it is negative at $N = N_0$; and as N gets very large, the sample-based SDP policy approaches the true optimal policy and gives a positive expected profit. Thus at this β value the expected profit as a function of sample size has a region where it decreases followed by a region where it increases. In the numerical experiments of the next section we will see that this pattern of a dip in SDP performance also occurs for other distributions. In fact the effect is pronounced for lognormal distributions which have more weight in their tails than exponential distributions.

To compute the expected out-of-sample performance of the sample-based policies on (ROP) for different price distributions, we use a simulation coded in the `Julia` programming language (Bezanson et al. 2017). Although (ROP) has an infinite horizon, simulation with a finite number of stages (say T) will provide accurate performance estimates as long as T is sufficiently large; we set $T \coloneqq 1 \times 10^3$. With $x^1 \coloneqq 1$, $\beta \coloneqq 0.99$, and $C(x) \coloneqq \frac{1}{2}x^2$, we

1. Sample N random prices $\tilde{p}_1, \dots, \tilde{p}_N \sim P$ to construct a sample-based policy y using equation (10) for SDP or equation (11) for MPC.
2. Sample a random price $\tilde{p}^t \sim P$, compute $\beta^{t-1}\big(\tilde{p}^t(x^t - y(x^t, \tilde{p}^t)) - C(y(x^t, \tilde{p}^t))\big)$, and set the next inventory level $x^{t+1} \coloneqq y(x^t, \tilde{p}^t)$.
3. Repeat Step 2 from stage $t = 1$ to stage $t = T$ to realize total discounted reward $\sum_{t=1}^{T} \beta^{t-1}\big(\tilde{p}^t(x^t - y(x^t, \tilde{p}^t)) - C(y(x^t, \tilde{p}^t))\big)$.

We repeat Steps 1 to 3 to generate random out-of-sample performance realizations. The average of these realizations can then be used as a statistical estimate of expected out-of-sample performance. In our simulations we use 1×10^5 realizations and found this sufficient to ensure accurate results—in Figures 1 to 5 the standard error intervals are smaller than the markers and so are not shown.

5.3. Skewed price distributions

Let Triangular(a, m, b) denote a triangular distribution with lower limit a, mode m, and upper limit b. Figures 1 and 2 present the expected out-of-sample performance of sample-based SDP and MPC on (ROP) for left- and right-skewed triangular price distributions with mean 1 and variance $^1/_8$.

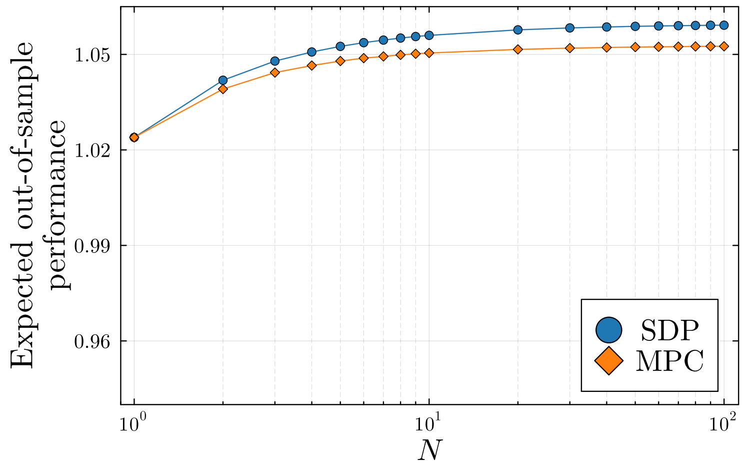

Figure 1 **Performance of sample-based SDP and MPC on (ROP) for Triangular$(0, {}^3/_2, {}^3/_2)$ distributed price.**

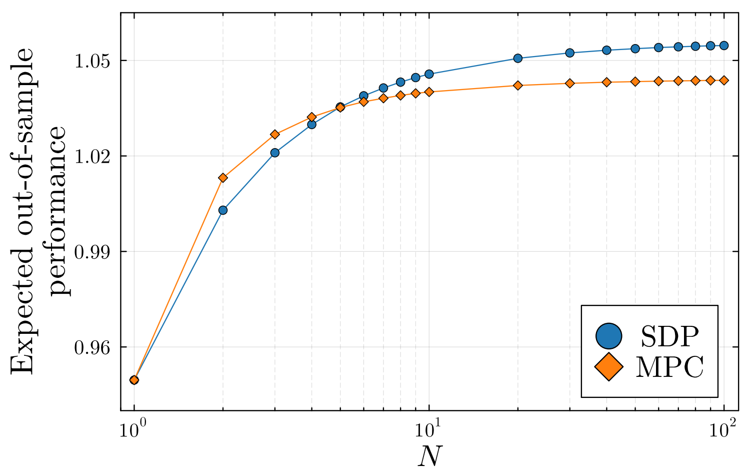

Figure 2 **Performance of sample-based SDP and MPC on on (ROP) for Triangular$({}^1/_2, {}^1/_2, 2)$ distributed price.**

Figure 1 shows SDP outperforming MPC for all N when the true price distribution is left-skewed. In contrast, Figure 2 shows MPC outperforming SDP for $N \leq 4$ when the true price distribution is right-skewed. The possibility of a particular realisation of the N price samples containing a price high enough to cause SDP to under-sell and overpay in storage costs is more likely in the presence of right-skew.

Figure 3 presents the expected out-of-sample performance of sample-based SDP and MPC on (ROP) for a right-skewed triangular price distribution with mean 1 and variance ${}^1/_2$.

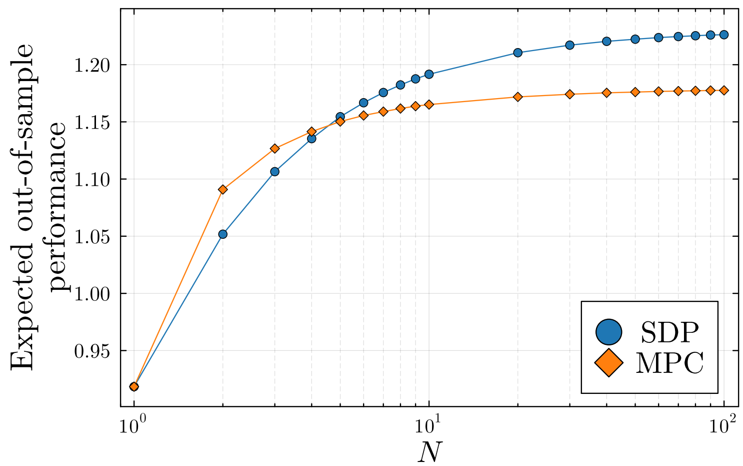

Figure 3 Performance of sample-based SDP and MPC on (ROP) for Triangular$(0,0,3)$ distributed price.

Figure 3 shows MPC outperforming SDP for $N \leq 4$ when the true price distribution is right-skewed and has a higher variance than the price distribution used in Figure 2. This is the same range of sample sizes as that in Figure 2, but the relative amount by which MPC outperforms SDP is increased.

5.4. Price distributions with a right-hand tail

Recall that Exponential(λ) is an exponential distribution with rate λ, and let LogNormal(μ, σ^2) denote a lognormal distribution with mean μ and variance σ^2. Figures 4 and 5 present the expected out-of-sample performance of sample-based SDP and MPC on (ROP) for exponential and lognormal price distributions with mean 1.

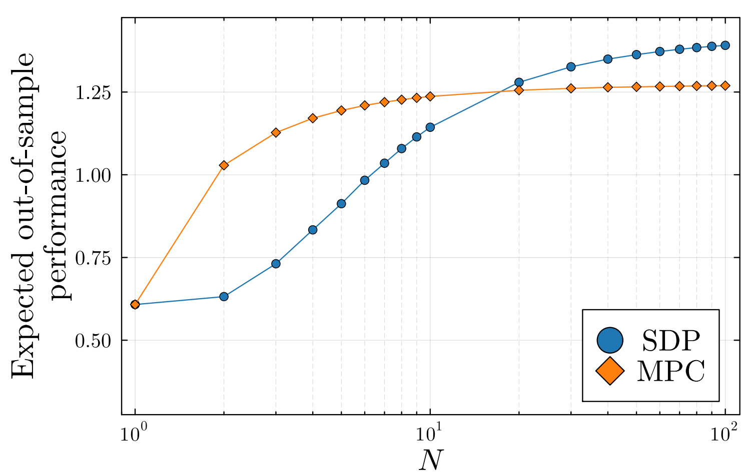

Figure 4 Performance of sample-based SDP and MPC on (ROP) for Exponential(1) distributed price.

Figure 4 shows MPC outperforming SDP for $N \leq 10$ when the true distribution is exponential. This is a larger range than that in Figures 1 to 3. The right-hand tail increases the propensity for very high prices to be included in the N samples, which worsens the relative performance of SDP.

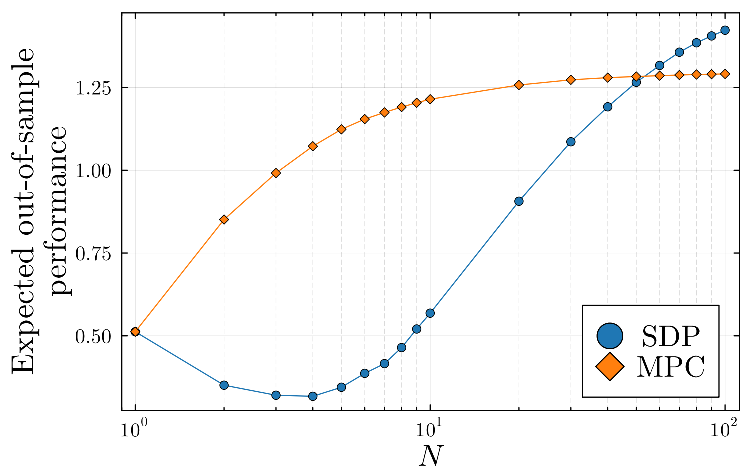

Figure 5 Performance of sample-based SDP and MPC on (ROP) for LogNormal$(-1/2, 1)$ distributed price.

Figure 5 shows MPC outperforming SDP for $N \leq 50$ when the true distribution is right-skewed and has a heavier tail than the exponential price distribution used in Figure 4. The poor performance of SDP exhibited in Figure 5 suggests that a result similar to that of Proposition 4 for exponential distributions may also hold for lognormal distributions. Considering that lognormal distributions have more weights in their tails than exponential distributions, this would not be surprising.

For $N = 2$, Figure 6 presents the relative out-of-sample performance of sample-based SDP and MPC on (ROP) as a function of the samples used to determine each policy.

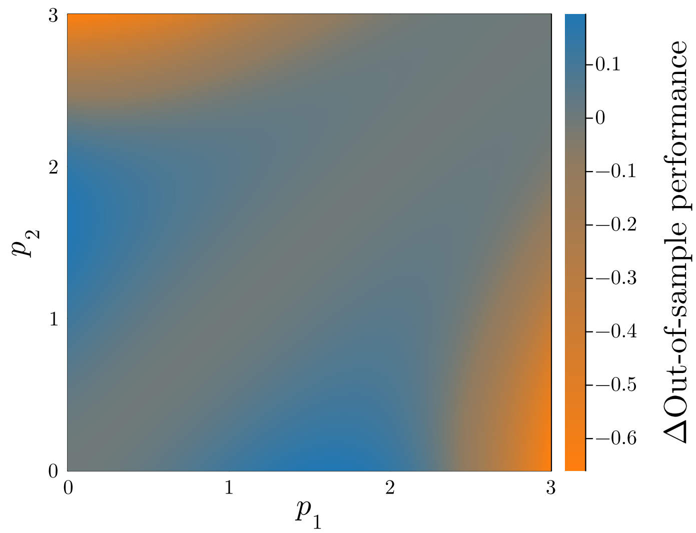

Figure 6 **Performance of sample-based SDP minus that of MPC on (ROP), as a function of two price samples for LogNormal($-1/2, 1$) distributed price. Orange regions indicate samples for which the MPC policy outperforms the SDP policy and blue regions indicate the opposite.**

Figure 6 explicitly demonstrates that a single very high price sample worsens the performance of SDP more than that of MPC; typical samples result in the SDP policy outperforming the MPC policy, but for extreme events where one price sample is very high the reverse occurs and the MPC policy outperforms the SDP policy.

6. Discussion

In this paper we provided an explanation for the good out-of-sample performance of MPC that is sometimes observed in practice, based on the interpretation of MPC as solving a mean-constrained distributionally-robust optimization problem. This depends critically on the MPC Bellman operators associated with the problem being concavity preserving. When the operator is instead convexity preserving, the interpretation becomes distributionally optimistic. Loosely speaking, the problem type for which the robust interpretation holds has uncertainty in the objective function parameters (as in Downward et al. 2020), and that for which the optimistic interpretation holds features constraint uncertainty (as in Pereira and Pinto 1991). These features provide some guidance on when MPC may-or-may-not be the right choice to address uncertainty in a particular problem. However, this is not the whole story, since out-of-sample performance also depends on the form of the underlying probability distributions present in a particular problem.

We then studied the performance of sample-based SDP and MPC on a multistage stochastic revenue optimization problem of the type in Downward et al. (2020), finding that SDP can be outperformed by MPC when the underlying price distribution is right-skewed or has a large right-hand tail and the number of samples in the distributional approximations is not too large. In the case where the underlying price distribution is exponential and the discount factor approaches 1, SDP can be outperformed by MPC regardless of the number of samples.

The performance issues which occur when applying sample-based SDP to the stochastic revenue optimization problem may be alleviated by appending newly-observed prices to the sample history and updating the policy before applying it again. In general, this approach is not practical as the time complexities of algorithms used to solve for the optimal (in-sample) SDP policy grow quickly in the number of samples. It could be more effective to use some sort of rolling window so that policies are based on a subset of the most recent samples. Still, this requires recalculation of the optimal policy at each stage, which may not be desirable. Also notice that whenever the rolling window includes a very high price sample, the same performance issues will occur.

The stochastic revenue optimization problem we have considered is quite restricted; having deterministic dynamics and constraints that do not allow inventory to increase. This gives a transient problem of selling inventory down rather than a problem with a steady-state component. We have chosen this problem because it enables an analytical solution and a more detailed analysis, but we expect that similar results would be obtained for problems in which there are occasional additional amounts of inventory arriving. For example, if stock to replenish inventory to the initial level x^1 arrives in each time period with probability α, then a renewal theory argument shows that the problem of maximizing average reward per unit time is equivalent to the revenue optimization problem with the discount factor $\beta = 1 - \alpha$.

Appendix A: Proof of Proposition 1

Proposition 1. *Let Assumption 1 hold. Suppose that the set-valued function $\mathcal{Y}(x)$ does not depend on ξ and that the map $\xi \mapsto \varphi(x,y,\xi)$ is concave for each $x,y \in \mathcal{X}$. Then the MPC Bellman operator B_{M} is concavity preserving.*

Proof of Proposition 1 Consider the set

$$\mathcal{F} := \left\{ f \in \mathcal{C}(\mathcal{X} \times \Xi) \,\middle|\, \xi \mapsto f(x,\xi) \text{ is concave for each } x \in \mathcal{X} \right\}.$$

By definition, the map $\xi \mapsto f(x,\xi)$ is concave for each $x \in \mathcal{X}$ and $f \in \mathcal{F}$. It remains to be shown that for each $f \in \mathcal{F}$ we have $B_{\mathrm{M}}(f) \in \mathcal{F}$. Let $f \in \mathcal{F}$ and observe that for each $x \in \mathcal{X}$, $\xi, \xi' \in \Xi$, and $\lambda \in [0,1]$, it holds that

$$\begin{aligned}
B_{\mathrm{M}}(f)(x, \lambda\xi + (1-\lambda)\xi') &= \inf_{y \in \mathcal{Y}(x)} \left\{ \varphi(x,y,\lambda\xi + (1-\lambda)\xi') + \beta f(y,\mu_N) \right\} \\
&= \varphi(x,y^*,\lambda\xi + (1-\lambda)\xi') + \beta f(y^*,\mu_N) \\
&\geq \lambda\big(\varphi(x,y^*,\xi) + \beta f(y^*,\mu_N)\big) + (1-\lambda)\big(\varphi(x,y^*,\xi') + \beta f(y^*,\mu_N)\big) \\
&\geq \lambda \inf_{y \in \mathcal{Y}(x)} \left\{ \varphi(x,y,\xi) + \beta f(y,\mu_N) \right\} + (1-\lambda) \inf_{y \in \mathcal{Y}(x)} \left\{ \varphi(x,y,\xi') + \beta f(y,\mu_N) \right\} \\
&= \lambda B_{\mathrm{M}}(f)(x,\xi) + (1-\lambda) B_{\mathrm{M}}(f)(x,\xi'),
\end{aligned}$$

where the second equality holds for some $y^* \in \mathcal{Y}(x)$ since the infimum of a continuous function over a compact set is attained, the first inequality holds due to the concavity of the map $\xi \mapsto \varphi(x,y^*,\xi)$, and the last inequality holds since $y^* \in \mathcal{Y}(x)$. This implies that $\xi \mapsto B_{\mathrm{M}}(f)(x,\xi)$ is concave for each $x \in \mathcal{X}$ and hence that $B_{\mathrm{M}}(f) \in \mathcal{F}$. $\square$

Appendix B: Proof of Theorem 1 and Corollary 1

Theorem 1. *Let Assumption 1 hold. For an ambiguity set $\mathcal{P}$, suppose that $\mathbb{E}_Q[\tilde{\xi}] = \mu_N$ for each $Q \in \mathcal{P}$ and that $\delta_{\mu_N} \in \mathcal{P}$. If the MPC Bellman operator B_{M} is concavity preserving, then the DRO functional equation (4) has a solution which is the same as the solution to the MPC functional equation (3).*

Proof of Theorem 1 The equation (4) features a DRO Bellman operator B_{R} which for $f \in \mathcal{C}(\mathcal{X} \times \Xi)$ at $(x,\xi) \in \mathcal{X} \times \Xi$ has value

$$B_{\mathrm{R}}(f)(x,\xi) := \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \sup_{Q \in \mathcal{P}} \mathbb{E}_P\big[f(y,\tilde{\xi})\big] \Big\}.$$

With B_{M} concavity preserving, there exists a closed subset $\mathcal{F} \subset \mathcal{C}(\mathcal{X} \times \Xi)$ where the map $\xi \mapsto f(x,\xi)$ is concave for each $x \in \mathcal{X}$ and $f \in \mathcal{F}$, and which maps to itself under B_{M}. We will first show that B_{R} maps $\mathcal{F}$ into itself and that it is a contraction mapping on $\mathcal{F}$. To do this we will demonstrate that B_{R} agrees with B_{M} on $\mathcal{F}$.

Let $y \in \mathcal{X}$. For each $f \in \mathcal{F}$ we have that

$$\mathbb{E}_Q\big[f(y,\tilde{\xi})\big] \le f(y, \mathbb{E}_Q[\tilde{\xi}]) = f(y,\mu_N) \quad \forall Q \in \mathcal{P}, \tag{14}$$

where the inequality follows from Jensen's inequality (due to concavity), and the equality follows from the fact that $\mathbb{E}_Q[\tilde{\xi}] = \mu_N$ for all $Q \in \mathcal{P}$. The statement (14) implies that $\sup_{Q \in \mathcal{P}} \mathbb{E}_Q\big[f(y,\tilde{\xi})\big] \le f(y,\mu_N)$. But, with $\mathbb{E}_{\delta_{\mu_N}}\big[f(y,\tilde{\xi})\big] = f(y,\mu_N)$ and $\delta_{\mu_N} \in \mathcal{P}$, the supremum is attained by δ_{μ_N} and $\sup_{Q \in \mathcal{P}} \mathbb{E}_Q\big[f(y,\tilde{\xi})\big] = f(y,\mu_N)$. Thus, for each $f \in \mathcal{F}$ we have that

$$\begin{aligned} B_{\mathrm{R}}(f)(x,\xi) &= \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \sup_{Q \in \mathcal{P}} \mathbb{E}_Q\big[f(y,\tilde{\xi})\big] \Big\} \\ &= \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta f(y,\mu_N) \Big\} = B_{\mathrm{M}}(f)(x,\xi) \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \end{aligned} \tag{15}$$

Hence, B_{R} agrees with B_{M} on $\mathcal{F}$ under the sup norm. With B_{M} mapping $\mathcal{F}$ to itself and being a contraction mapping on $\mathcal{F}$ under the sup norm, the same holds for B_{R}.

It remains to show that the DRO functional equation (4) has a solution which is the same as the solution to the MPC functional equation (3). Observe that $(\mathcal{F}, \|\cdot\|_\infty)$ is a Banach space as $\mathcal{F}$ is closed. Since B_{R} is a contraction mapping on $\mathcal{F}$, Banach's Fixed-Point Theorem shows that the equation $f = B_{\mathrm{R}}(f)$ has a unique solution in $\mathcal{F}$, and this solution solves the DRO functional equation (4). In light of (15), this solution is the same as the solution to (3). $\square$

Corollary 1. *Suppose that the conditions in the statement of Theorem 1 hold, but that the MPC Bellman operator B_{M} is convexity preserving rather than concavity preserving. Then the DOO functional equation (5) has a solution which is the same as the solution to the MPC functional equation (3).*

Proof of Corollary 1 The proof repeats the proof of Theorem 1, but with the inequalities reversed and $\inf_{Q \in \mathcal{P}} \mathbb{E}_Q\big[f(y,\tilde{\xi})\big]$ attained by δ_{μ_N}. $\square$

Appendix C: Proof of Theorem 2 and Corollary 2

THEOREM 2. *Let Assumption 1 hold, and set $\mu_N = \mu := \mathbb{E}_P[\tilde{\xi}]$ in the MPC functional equation (3). If the resulting MPC Bellman operator B_M is concavity preserving, then*

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \leq v_{\mathrm{M}}(x^1, \mu) \quad \textit{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N};$$

that is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (3) is an upper-bound on the cost incurred when applying the MPC policy to (SOC).

Proof Let $\mathcal{P} := \{P, \delta_\mu\}$, and recall that $\mathcal{P}^\infty := \mathcal{P} \times \mathcal{P} \times \cdots$ is a rectangular set. By Theorem 1, the MPC policy solves (DR-SSOC), and we have that

$$\begin{aligned} v_{\mathrm{M}}(x^1, \mu) &= \inf_{y^1, y^2, \ldots} \sup_{R \in \mathcal{P}^\infty} \mathbb{E}_R\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \quad \text{subject to } x^{t+1} = y^t(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N} \\ &= \sup_{R \in \mathcal{P}^\infty} \mathbb{E}_R\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \quad \text{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N}. \end{aligned} \tag{16}$$

The cost incurred when applying the MPC policy to (SOC) is

$$\mathbb{E}_{P^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \leq \sup_{R \in \mathcal{P}^\infty} \mathbb{E}_R\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \quad \text{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N},$$

which holds since $P^\infty \in \mathcal{P}^\infty$. The right-hand side equals $v_{\mathrm{M}}(x^1, \mu)$, by virtue of (16). □

COROLLARY 2. *Suppose that the conditions in the statement of Theorem 2 hold, but that the MPC Bellman operator is convexity preserving rather than concavity preserving. Then*

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1}\varphi(x^t, x^{t+1}, \tilde{\xi}^t)\right] \geq v_{\mathrm{M}}(x^1, \mu) \quad \textit{where } x^{t+1} = y_{\mathrm{M}}(x^t, \tilde{\xi}^t)\ \forall t \in \mathbb{N};$$

that is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (3) is a lower-bound on the cost incurred when applying the MPC policy to (SOC).

Proof of Corollary 2 The proof repeats the proof of Theorem 2, but with the reference to Theorem 1 replaced by a reference to Corollary 1, the supremum operations replaced by infimum operations, and the inequalities reversed. □

Appendix D: Proof of Theorem 3

We prove the theorem by induction. To do this we extend Definition 2 as follows.

DEFINITION 3. Given two policies $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ which are feasible for (SOC), let $\boldsymbol{V}^1_{yy'}(x) := \boldsymbol{V}_{yy'}(x)$, and for each $T \in \mathbb{N}$ define the function $\boldsymbol{V}^{T+1}_{yy'} : \mathcal{X} \mapsto \mathbb{R}$ by

$$\boldsymbol{V}^{T+1}_{yy'}(x) := \mathbb{E}\left[\varphi\big(x, y(x, \tilde{\xi}), \tilde{\xi}\big) + \beta \boldsymbol{V}^T_{yy'}\big(y(x, \tilde{\xi})\big)\right]. \tag{17}$$

The value $\boldsymbol{V}^T_{yy'}(x)$ is the objective value of (SOC) starting from initial state x if the policy y is used for the first T stages and then the policy y' is used thereafter. It is clear that $\boldsymbol{V}^T_{yy'}$ is well-defined and bounded in the same way that $\boldsymbol{V}_y$ and $\boldsymbol{V}_{y'}$ are.

LEMMA 1. *Let Assumption 2 hold. For two policies $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ that are feasible for (SOC), it holds that* $\lim_{T\to\infty} \left| \boldsymbol{V}^T_{yy'}(x) - \boldsymbol{V}_y(x) \right| = 0$ *for each $x \in \mathcal{X}$.*

Proof of Lemma 1. For each $x^1 \in \mathcal{X}$ observe that

$$\begin{aligned}\boldsymbol{V}^1_{yy'}(x^1) - \boldsymbol{V}_y(x^1) &= \mathbb{E}\left[\varphi\big(x^1, y(x^1,\tilde\xi), \tilde\xi\big) + \beta \boldsymbol{V}_{y'}\big(y(x^1,\tilde\xi)\big)\right] - \mathbb{E}\left[\varphi\big(x^1, y(x^1,\tilde\xi), \tilde\xi\big) + \beta \boldsymbol{V}_{y}\big(y(x^1,\tilde\xi)\big)\right] \\ &= \beta \mathbb{E}\left[\boldsymbol{V}_{y'}\big(y(x^1,\tilde\xi)\big) - \boldsymbol{V}_{y}\big(y(x^1,\tilde\xi)\big)\right].\end{aligned}$$

Continuing this reasoning, a simple induction shows that

$$\left|\boldsymbol{V}^T_{yy'}(x^1) - \boldsymbol{V}_y(x^1)\right| = \beta^T \left| \mathbb{E}\left[\boldsymbol{V}_{y'}(x^T) - \boldsymbol{V}_y(x^T)\right]\right| \quad \text{where } x^{t+1} = y(x^t, \tilde\xi^t)\ \forall t \in \{1, \dots, T-1\}. \tag{18}$$

Since $\boldsymbol{V}_y$ and $\boldsymbol{V}_{y'}$ are bounded under Assumption 2 and $\beta^T \to 0$ as $T \to \infty$, the value of (18) goes to 0 as $T \to \infty$. Replacing x^1 with x concludes the proof. □

THEOREM 3. *Let Assumption 2 hold. Suppose that $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$ are two feasible policies for (SOC). If*

$$\boldsymbol{V}_{y'}(x) \le \boldsymbol{V}_{yy'}(x) \quad \forall x \in \mathcal{X},$$

then $\boldsymbol{V}_{y'}(x) \le \boldsymbol{V}_y(x)$ for each $x \in \mathcal{X}$.

Proof of Theorem 3. Denoting $\boldsymbol{V}_{y'}(x)$ by $\boldsymbol{V}^0_{yy'}(x)$, we pose the inductive hypothesis $\boldsymbol{V}^{t-1}_{yy'}(x) \le \boldsymbol{V}^t_{yy'}(x)$ for all $x \in \mathcal{X}$, which is true for $t = 1$ by assumption. By the hypothesis, $\boldsymbol{V}^{t-1}_{yy'}(y(x,\xi)) \le \boldsymbol{V}^t_{yy'}(y(x,\xi))$, and it follows that

$$\begin{aligned}\boldsymbol{V}^t_{yy'}(x) &= \mathbb{E}\left[C\big(x, y(x,\tilde\xi), \tilde\xi\big) + \beta \boldsymbol{V}^{t-1}_{yy'}\big(y(x,\tilde\xi)\big)\right] \\ &\le \mathbb{E}\left[C\big(x, y(x,\tilde\xi), \tilde\xi\big) + \beta \boldsymbol{V}^{t}_{yy'}\big(y(x,\tilde\xi)\big)\right] = \boldsymbol{V}^{t+1}_{yy'}(x)\end{aligned} \tag{19}$$

for all $x \in \mathcal{X}$. The inequality (19) establishes the induction. Hence, $\boldsymbol{V}_{y'}(x) \le \boldsymbol{V}^t_{yy'}(x)$ for all $t \in \mathbb{N}$ and $\boldsymbol{V}_{y'}(x) \le \lim_{t\to\infty} \boldsymbol{V}^t_{yy'}(x) = \boldsymbol{V}_y(x)$, where Lemma 1 yields the final equality. □

Appendix E: Proof of Proposition 2

PROPOSITION 2. *Suppose that the probability distribution P has compact support. Then the policy $y : \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}_+$ which for inventory x and price p sells down to the inventory level*

$$y(x,p) = c^{-1}\left(\Big(\beta \mathbb{E}_P\left[(\tilde p - p)_+\right] - (1-\beta)p\Big)_{[c(0), c(x)]}\right)$$

is optimal for (ROP).

Proof of Proposition 2. Recall that when P has compact support there is a unique continuous function $v : \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}$ which solves equation (8), and that the mapping $x \mapsto v(x,p)$ is concave for each $p \in \mathbb{R}$. Given v, we are going to derive a closed-form expression for a feasible policy satisfying (9), which therefore solves (ROP). Define the function $V : \mathbb{R}_+ \mapsto \mathbb{R}$ by $V(x) := \mathbb{E}\big[v(x,\tilde p)\big]$ and note that V is continuous (Shapiro et al.

2009, Theorem 7.43) and concave (Shapiro et al. 2009, Theorem 7.46). Satisfying (9) amounts to solving for each $(x,p)\in\mathbb{R}_+\times\mathbb{R}$ the optimization

$$\begin{aligned} \underset{y}{\text{maximize}} \quad & \phi_p(y) := p(x-y) - C(y) + \beta V(y) \\ \text{subject to} \quad & 0 \le y \le x \\ & y \in \mathbb{R}_+. \end{aligned} \tag{20}$$

The function $\phi_p:\mathbb{R}_+\mapsto\mathbb{R}$ is continuous and strictly concave so the optimization (20) which is over a compact set has a unique solution $y^*(x,p)\in[0,x]$. We let $y^*(\infty,p) := \lim_{x\to\infty} y^*(x',p)$ and note that this is a finite value since the map $y\mapsto\phi_p(y)$ is eventually decreasing as $\lim_{y\to\infty} -c(y) = -\infty$. Observe that $y^*(\infty,p)$ is equal to $y^*(x,p)$ when projected onto the compact interval $[0,x]$, and that the mapping $p\mapsto y^*(\infty,p)$ is decreasing. Hence, there exists a $p^*(x)\in[-\infty,\infty]$ such that, for each $p\in\mathbb{R}$, if $p^*(x)\le p$, then $x\ge y(\infty,p)$, and if $p\le p^*(x)$, then $y(\infty,p)\ge x$. Comparing $p^*(x)$ and p then characterises whether-or-not $y^*(\infty,p)$ needs to be projected onto $[0,x]$ to solve (20).

We will proceed by deriving an expression for the derivative of V using a fixed-point equation. In fact we will need to work with the superdifferential ∂ operator since V may be nonsmooth. With V continuous and concave the superdifferential $\partial V(x)$ is nonempty and compact, and we denote by $\partial v(x,p)$ the superdifferential (with respect to x) of v at (x,p). When $p^*(x)\le p$, by definition $x\ge y^*(\infty,p)$, and so $y^*(\infty,p)=y^*(x,p)$. Together with (8) this implies that

$$\begin{aligned} v(x,p) = p\big(x-y^*(\infty,p)\big) - C\big(y^*(\infty,p)\big) + \beta V\big(y^*(\infty,p)\big) \quad \forall p\ge p^*(x) \\ \implies p\in\partial v(x,p) \quad \forall p\ge p^*(x). \end{aligned} \tag{21}$$

Conversely,

$$\begin{aligned} v(x,p) = -C(x) + \beta V(x) \quad \forall p\le p^*(x) \\ \implies -c(x)+\beta\partial V(x)\subset\partial v(x,p) \quad \forall p\le p^*(x). \end{aligned} \tag{22}$$

Define the function $\psi_x:\partial V(x)\times\mathbb{R}\mapsto\mathbb{R}$ by

$$\psi_x(g,p) := \begin{cases} p & \text{if } p\ge p^*(x) \\ -c(x)+\beta g & \text{if } p< p^*(x). \end{cases} \tag{23}$$

Observe that the first case of (23) coincides with (21) and the second case of (23) coincides with (22), so, for each $g\in\partial V(x)$ we have that

$$\psi_x(g,p)\in\partial v(x,p) \quad \forall p\in\mathbb{R},$$

and thus $\mathbb{E}\big[\psi_x(g,\tilde p)\big]\in\partial V(x)$ (Shapiro et al. 2009, Theorem 7.47). Substituting (23) into the expectation operator shows that

$$\mathbb{E}\big[\psi_x(g,\tilde p)\big] = \mathbb{P}\big[\tilde p < p^*(x)\big]\big(\beta g - c(x)\big) + \mathbb{P}\big[\tilde p\ge p^*(x)\big]\mathbb{E}\big[\tilde p \,\big|\, \tilde p\ge p^*(x)\big],$$

and for each $g,g'\in\partial V(x)$

$$\Big|\mathbb{E}\big[\psi_x(g,\tilde p)\big] - \mathbb{E}\big[\psi_x(g',\tilde p)\big]\Big| = \mathbb{P}\big[\tilde p < p^*(x)\big]\beta|g-g'| < |g-g'|. \tag{24}$$

What we have just shown is that $g \mapsto \mathbb{E}\big[\psi_x(g,\tilde{p})\big]$ is a contraction mapping. As $\partial V(x) \subset \mathbb{R}$ is a nonempty and compact set, $(\partial V(x), |\cdot|)$ is a complete metric space, and by the Banach Fixed-Point Theorem there is a unique $g(x) \in \partial V(x)$ satisfying $g(x) = \mathbb{E}\big[\psi_x(g(x),\tilde{p})\big]$. The equation (24) yields

$$\begin{aligned} g(x) &= \mathbb{P}\big[\tilde{p} < p^*(x)\big]\big(\beta g(x) - c(x)\big) + \mathbb{P}\big[\tilde{p} \geq p^*(x)\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} \geq p^*(x)\big] \\ &= \frac{\mathbb{P}\big[\tilde{p} \geq p^*(x)\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} \geq p^*(x)\big] - \mathbb{P}\big[\tilde{p} < p^*(x)\big]c(x)}{1 - \beta\mathbb{P}\big[\tilde{p} < p^*(x)\big]} \in \partial V(x). \end{aligned} \tag{25}$$

Using the expression (25) and observing that p satisfies the definition of $p^*(y(\infty,p))$, a first-order optimality condition for the unconstrained version of (20) is

$$\begin{aligned} 0 &= -p - c(y) + \beta\frac{\mathbb{P}\big[\tilde{p} \geq p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} \geq p\big] - \mathbb{P}\big[\tilde{p} < p\big]c(y)}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]} \\ &= -p - \frac{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]}c(y) + \beta\frac{\mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big]}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]} - \frac{\beta\mathbb{P}\big[\tilde{p} < p\big]}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]}c(y) \\ &= -p - \frac{c(y)}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]} + \beta\frac{\mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big]}{1 - \beta\mathbb{P}\big[\tilde{p} < p\big]} \\ &= -p\big(1 - \beta\mathbb{P}\big[\tilde{p} < p\big]\big) - c(y) + \beta\mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big] \end{aligned}$$

which can be re-arranged to

$$\begin{aligned} c(y) &= -p\big(1 - \beta\mathbb{P}\big[\tilde{p} < p\big]\big) + \beta\mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big] \\ &= \beta p\mathbb{P}\big[\tilde{p} < p\big] + \beta\mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big] - p \\ &= \beta\big(p\mathbb{P}\big[\tilde{p} < p\big] + \mathbb{P}\big[\tilde{p} > p\big]\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} > p\big]\big) - p \\ &= \beta\big(\mathbb{E}\big[(\tilde{p} - p)_+\big] + p\big) - p \\ &= \beta\mathbb{E}\big[(\tilde{p} - p)_+\big] - (1 - \beta)p. \end{aligned} \tag{26}$$

If $\mathbb{E}\big[(\tilde{p} - p)_+\big] - (1-\beta)p \in [c(0), c(x)]$, then (26) shows that

$$c^{-1}\Big(\beta\mathbb{E}\big[(\tilde{p} - p)_+\big] + \beta p - p\Big)$$

solves (20). Otherwise, if $\mathbb{E}\big[(\tilde{p}-p)_+\big] - (1-\beta)p > c(x)$, then x solves (20). Similarly, if $\mathbb{E}\big[(\tilde{p}-p)_+\big] - (1-\beta)p < c(0)$, then 0 solves (20). Combining these three cases yields

$$y^*(x,p) = c^{-1}\bigg(\Big(\beta\mathbb{E}\big[(\tilde{p} - p)_+\big] + \beta p - p\Big)_{[c(0),c(x)]}\bigg).$$

□

Appendix F: Proof of Proposition 3

We prove the result using several lemmas which provide formulae for $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{S}}(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{M}}(x)$. The following result will be helpful.

LEMMA 2. *(Shapiro et al. 2009, Theorem 7.44) Let $f : \mathbb{R}^m \times \Omega \to \mathbb{R}$ be a random function with expected value $F(x) := \mathbb{E}\big[f(x,\tilde{\omega})\big]$. If the following conditions hold:*

(i) *The expectation $F(\bar{x})$ is well defined and finite valued at a given point $\bar{x} \in \mathbb{R}^m$,*

(ii) *There exists a positive-valued random variable* $L(\tilde{\omega})$ *such that* $\mathbb{E}\big[L(\tilde{\omega})\big] < \infty$*, and for each* x, x' *in a neighbourhood of* $\bar{x}$ *and almost every* $\omega \in \Omega$ *the following inequaltiy holds:*

$$|f(x,\omega) - f(x',\omega)| \le L(\omega)\|x - x'\|_2,$$

(iii) *For almost every* $\omega \in \Omega$ *the function* $f(x,\omega)$ *is differentiable with respect to* x *at* $\bar{x}$*,*
Then $F(x)$ *is differentiable at* $\bar{x}$ *and*

$$\nabla F(\bar{x}) = \mathbb{E}\big[\nabla_x f(\bar{x}, \tilde{\omega})\big].$$

Now the derivative values can be established.

Lemma 3. *Let Assumption 3 hold. Then the functions* $\boldsymbol{V}_{\mathrm{S}}$ *and* $\boldsymbol{V}_{\mathrm{M}}$ *are differentiable.*

Proof of Lemma 3. We will proceed by verifying that the conditions of Lemma 2 apply to $\boldsymbol{V}_{\mathrm{S}}$. Under Assumption 3, note that $\boldsymbol{V}_{\mathrm{S}}$ is well-defined and finite valued, satisfying condition (i) of Lemma 2. Next we address condition (iii). Define the function $\boldsymbol{v}_{\mathrm{S}} : \mathbb{R}_+ \times \mathbb{R}_+^\infty \mapsto \mathbb{R}$ by

$$\boldsymbol{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots) := \sum_{t=1}^{\infty} \beta^{t-1}\big(p^t(x^t - x^{t+1}) - C(x^{t+1})\big) \quad \text{where } x^{t+1} = y_{\mathrm{S}}(x^t, p^t)\ \forall t \in \mathbb{N}. \tag{27}$$

If $\tilde{p}^1, \tilde{p}^2, \ldots$ are random price variables that are independent and identically distributed according to P, the definition (6) shows that $\mathbb{E}\big[\boldsymbol{v}_{\mathrm{S}}(x^1, \tilde{p}^1, \tilde{p}^2, \ldots)\big] = \boldsymbol{V}_{\mathrm{S}}(x^1)$. Recall that $p^*_{\mathrm{S}}(x^1)$ is the minimum price required for the SDP policy to decide that it is worthwhile to sell a portion of inventory from x^1, and that for any $p \ge p^*_{\mathrm{S}}(x^1)$ the SDP policy sells down to the target inventory level $y_{\mathrm{S}}(\infty, p) \le x^1$. It follows that in (27) all inventory levels $x^t = x^1$ for $t \le T$, with T being the first time at which $p^T \ge p^*_{\mathrm{S}}(x^1)$. The next inventory level $x^{T+1} = y_{\mathrm{S}}(\infty, p^T)$, and along with the prices $p^{T+1}, p^{T+2}, \ldots$, these uniquely determine the remaining inventory levels $x^{T+2}, x^{T+3}, \ldots$ when following the sample-based SDP policy. Under Assumption 3 the probability distribution P is atomless which implies that $\mathbb{P}\big[\tilde{p}^t \ne p^*_{\mathrm{S}}(x^1)\big] = 1$. It follows that in (27) there are no prices p^t with $y_{\mathrm{S}}(\infty, p^t) = x^1$ almost surely. Hence, there is a neighbourhood of x^1 values in which $x^t = x^1$ for $t \le T$ and $x^{T+1} = y_{\mathrm{S}}(\infty, p^T)$. The remaining inventory values $x^{T+2}, x^{T+3}, \ldots$ are fixed within this neighbourhood. Thus, we may write

$$\frac{\mathrm{d}\boldsymbol{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots)}{\mathrm{d}x^1} = \beta^{T-1} p^T - \sum_{t=1}^{T-1} \beta^{t-1} c(x^1), \tag{28}$$

meeting condition (iii) of Lemma 2.

It remains to verify condition (ii) of Lemma 2. When attempting to evaluate $\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}x^1}\boldsymbol{v}_{\mathrm{S}}(x^1, \tilde{p}^1, \tilde{p}^2, \ldots)\right]$ using the expression (28), the time T and the price p^T are random variables. The expectation of the first term in (28) is $\mathbb{E}\big[\beta^{T-1}p^T\big] \le \mathbb{E}\big[\tilde{p} \,\big|\, \tilde{p} \ge p^*_{\mathrm{S}}(x^1)\big] \le \mathbb{E}\big[\tilde{p} \,\big|\, \tilde{p} \ge \max\{p_1, \ldots, p_N\}\big] < \infty$, where the first inequality follows since $p^T \ge p^*_{\mathrm{S}}(x^1)$, the second inequality follows since the SDP policy always decides that it is worthwhile to sell when the price is higher than the highest price sample, and the third inequality follows since P has a finite mean. The expectation of the second term in (28) is $\mathbb{E}\big[\sum_{t=1}^{\tilde{T}-1} \beta^{t-1} c(x^1)\big] \le \frac{1}{1-\beta} c(x^1)$ which is bounded on compact sets. Together these observations show that $\boldsymbol{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots)$ has a Lipschitz constant in its first argument with finite expectation, meeting condition (ii) of Lemma 2. Thus, Lemma 2 applies and $\boldsymbol{V}_{\mathrm{S}}(x)$ is differentiable. Similar reasoning shows that $\boldsymbol{V}_{\mathrm{M}}(x)$ is differentiable. □

Lemma 4. *Let Assumption 3 hold. Then the derivatives of the functions* $\boldsymbol{V}_\mathrm{S}$ *and* $\boldsymbol{V}_\mathrm{M}$ *are given by*

$$\frac{\mathrm{d}\boldsymbol{V}_\mathrm{S}(x)}{\mathrm{d}x} = \frac{\mathbb{E}_P\big[\tilde{p}\,\big|\,\tilde{p} \ge p_\mathrm{S}^*(x)\big]\mathbb{P}_P\big[\tilde{p} \ge p_\mathrm{S}^*(x)\big] - c(x)\mathbb{P}_P\big[\tilde{p} < p_\mathrm{S}^*(x)\big]}{1 - \beta\mathbb{P}_P\big[\tilde{p} < p_\mathrm{S}^*(x)\big]}$$

and

$$\frac{\mathrm{d}\boldsymbol{V}_\mathrm{M}(x)}{\mathrm{d}x} = \frac{\mathbb{E}_P\big[\tilde{p}\,\big|\,\tilde{p} \ge p_\mathrm{M}^*(x)\big]\mathbb{P}_P\big[\tilde{p} \ge p_\mathrm{M}^*(x)\big] - c(x)\mathbb{P}_P\big[\tilde{p} < p_\mathrm{M}^*(x)\big]}{1 - \beta\mathbb{P}_P\big[\tilde{p} < p_\mathrm{M}^*(x)\big]},$$

respectively.

Proof of Lemma 4. We proceed by showing that the derivative $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_\mathrm{S}(x)$ satisfies a functional equation. Define the function $\boldsymbol{v}_\mathrm{S} : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}$ by

$$\boldsymbol{v}_\mathrm{S}(x,p) := \begin{cases} -C(x) + \beta\boldsymbol{V}_\mathrm{S}(x) & \text{if } p < p_\mathrm{S}^*(x) \\ p\big(x - y_\mathrm{S}(\infty,p)\big) - C\big(y_\mathrm{S}(\infty,p)\big) + \beta\boldsymbol{V}_\mathrm{S}\big(y_\mathrm{S}(\infty,p)\big) & \text{if } p \ge p_\mathrm{S}^*(x). \end{cases} \tag{29}$$

The equation (12) shows that $\mathbb{E}_P\big[\boldsymbol{v}_\mathrm{S}(x,\tilde{p})\big] = \boldsymbol{V}_\mathrm{S}(x)$. Recalling that the map $x \mapsto p_\mathrm{S}^*(x)$ is continuous, for each $p < p_\mathrm{S}^*(x)$ there is a neighbourhood of x values in which the first case of (29) applies. Lemma 3 shows that $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_\mathrm{S}(x)$ exists and for each $p < p_\mathrm{S}^*(x)$ it follows that

$$\frac{\mathrm{d}\boldsymbol{v}_\mathrm{S}(x,p)}{\mathrm{d}x} = -c(x) + \beta\frac{\mathrm{d}\boldsymbol{V}_\mathrm{S}(x)}{\mathrm{d}x}. \tag{30}$$

Similarly, for each $p > p_\mathrm{S}^*(x)$ the second case of (29) applies and it follows that

$$\frac{\mathrm{d}\boldsymbol{v}_\mathrm{S}(x,p)}{\mathrm{d}x} = p. \tag{31}$$

Reasoning similar to that in the proof of Lemma 3 regarding Lemma 2 shows $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_\mathrm{S}(x) = \mathbb{E}\big[\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{v}_\mathrm{S}(x,\tilde{p})\big]$. Under Assumption 3, the probability distribution P is atomless which implies that $\mathbb{P}\big[\tilde{p} \ne p_\mathrm{S}^*(x)\big] = 1$; i.e. the event $\tilde{p} = p_\mathrm{S}^*(x)$ is immaterial when evaluating $\mathbb{E}\big[\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{v}_\mathrm{S}(x,\tilde{p})\big]$. Substituting the expressions (30) and (31) into the expectation yields

$$\frac{\mathrm{d}\boldsymbol{V}_\mathrm{S}(x)}{\mathrm{d}x} = \left(\beta\frac{\mathrm{d}\boldsymbol{V}_\mathrm{S}(x)}{\mathrm{d}x} - c(x)\right)\mathbb{P}\big[\tilde{p} < p_\mathrm{S}^*(x)\big] + \mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} \ge p_\mathrm{S}^*(x)\big]\mathbb{P}\big[\tilde{p} \ge p_\mathrm{S}^*(x)\big]$$

and rearranging gives the required expression

$$\frac{\mathrm{d}\boldsymbol{V}_\mathrm{S}(x)}{\mathrm{d}x} = \frac{\mathbb{E}\big[\tilde{p}\,\big|\,\tilde{p} \ge p_\mathrm{S}^*(x)\big]\mathbb{P}\big[\tilde{p} \ge p_\mathrm{S}^*(x)\big] - c(x)\mathbb{P}\big[\tilde{p} < p_\mathrm{S}^*(x)\big]}{1 - \beta\mathbb{P}\big[\tilde{p} < p_\mathrm{S}^*(x)\big]}.$$

The required expression for $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_\mathrm{M}(x)$ can be derived similarly. □

Now Proposition 3 can be established.

Proposition 3. *Let Assumption 3 hold and suppose that* P *has a probability density function* $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$*. For* N *price samples* $p_1, \ldots, p_N$ *which determine the sample-based SDP and MPC policies by (10) and (11), respectively, if*

$$c(x) \ge \beta\int_{p_\mathrm{S}^*(x)}^{\infty} pf(p)\mathrm{d}p \quad \forall x \in [0, x^1],$$

then $\boldsymbol{V}_\mathrm{M}(x) \ge \boldsymbol{V}_\mathrm{S}(x)$ *for each* $x \in [0, x^1]$*. That is, the MPC policy performs at least as well out-of-sample on (ROP) as the SDP policy.*

Proof of Proposition 3. For initial inventory $x \in [0, x^1]$, let $\boldsymbol{V}_{\mathrm{S,M}}(x)$ denote the objective value of (ROP) when the SDP policy is used for the first stage and the MPC policy is used thereafter (as in Definition 2). We will first show that $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{M}}(x) \geq \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{S,M}}(x)$ for each $x \in [0, x^1]$. Rearranging the identity given by Lemma 4 yields

$$\frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} = \left(\beta \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - c(x)\right) \int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p + \int_{p^*_{\mathrm{M}}(x)}^{\infty} pf(p)\mathrm{d}p.$$

Similarly,

$$\frac{\mathrm{d}\boldsymbol{V}_{\mathrm{S,M}}(x)}{\mathrm{d}x} = \left(\beta \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - c(x)\right) \int_{-\infty}^{p^*_{\mathrm{S}}(x)} f(p)\mathrm{d}p + \int_{p^*_{\mathrm{S}}(x)}^{\infty} pf(p)\mathrm{d}p,$$

and hence

$$\frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{S,M}}(x)}{\mathrm{d}x} = \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p - \left(\beta \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - c(x)\right) \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} f(p)\mathrm{d}p. \tag{32}$$

Again using the identity given by Lemma 4, the term $\beta \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{M}}(x) - c(x)$ in (32) can be written as

$$\begin{aligned}
\beta \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - c(x) &= \beta\left(\int_{p^*_{\mathrm{M}}(x)}^{\infty} pf(p)\mathrm{d}p - c(x)\int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p\right) \Big/ \left(1 - \beta \int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p\right) - c(x) \\
&= \left(\beta \int_{p^*_{\mathrm{M}}(x)}^{\infty} pf(p)\mathrm{d}p - c(x)\right) \Big/ \left(1 - \beta \int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p\right) \\
&\leq \left(\beta \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p\right) \Big/ \left(1 - \beta \int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p\right),
\end{aligned} \tag{33}$$

where the inequality follows from the condition in the statement of the proposition. Recall $p^*_{\mathrm{S}}(x) \geq p^*_{\mathrm{M}}(x)$ and assume $p^*_{\mathrm{S}}(x) \neq p^*_{\mathrm{M}}(x)$ (otherwise (32) shows that $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{M}}(x) \geq \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{S,M}}(x)$ holds trivially). This implies that $p^*_{\mathrm{S}}(x) > p^*_{\mathrm{M}}(x)$. Thus,

$$\left(\beta \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p\right) \Big/ \left(1 - \beta \int_{-\infty}^{p^*_{\mathrm{M}}(x)} f(p)\mathrm{d}p\right) \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} f(p)\mathrm{d}p \leq \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p, \tag{34}$$

since $\int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p > 0$ can be cancelled from both sides and the identity rearranged to $\beta \int_{-\infty}^{p^*_{\mathrm{S}}(x)} f(p)\mathrm{d}p \leq 1$ which is clearly true. In view of (33) and (34), it holds that

$$\left(\beta \frac{\mathrm{d}\boldsymbol{V}_{\mathrm{M}}(x)}{\mathrm{d}x} - c(x)\right) \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} f(p)\mathrm{d}p \leq \int_{p^*_{\mathrm{M}}(x)}^{p^*_{\mathrm{S}}(x)} pf(p)\mathrm{d}p,$$

which shows that (32) is nonnegative, whereby $\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{M}}(x) \geq \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{S,M}}(x)$ for each $x \in [0, x^1]$ and so $\boldsymbol{V}_{\mathrm{M}}(x) \geq \boldsymbol{V}_{\mathrm{S,M}}(x)$ for each $x \in [0, x^1]$. Applying Theorem 3, we have $\boldsymbol{V}_{\mathrm{M}}(x) \geq \boldsymbol{V}_{\mathrm{S}}(x)$, as required. □

Appendix G: Proof of Proposition 4

To make the dependence on the samples explicit in this appendix, for initial inventory level x and N price samples $p_1, \ldots, p_N$ we write $\boldsymbol{V}_{\mathrm{S}}(x; p_1, \ldots, p_N)$ and $\boldsymbol{V}_{\mathrm{M}}(x; p_1, \ldots, p_N)$ to denote the out-of-sample performance of sample-based SDP and MPC, respectively, on (ROP). If $\tilde{p}_1, \ldots, \tilde{p}_N$ are independent and identically distributed according to P, then $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{S}}(x; \tilde{p}_1, \ldots, \tilde{p}_N)\big]$ and $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{M}}(x; \tilde{p}_1, \ldots, \tilde{p}_N)\big]$ is the expected out-of-sample performance of sample-based SDP and MPC, respectively, on (ROP).

Proposition 4 *Let P be an* Exponential(1) *distribution, $x^1 := 1$, and $C(x) := \frac{1}{2}x^2$. For each sample size $N \geq 2$, as the discount factor $\beta \to 1$ the expected out-of-sample performance of sample-based SDP on (ROP) is unbounded below while the expected out-of-sample performance of sample-based MPC on (ROP) is bounded.*

Proof of Proposition 4 The Exponential(1) distribution has nonnegative support and for $p \geq 0$ has cumulative distribution function $F(p) = 1 - e^{-p}$ and probability density function $f(p) = e^{-p}$. Note that Assumption 3 applies here. Also, with $C(x) = \frac{1}{2}x^2$ the derivative function $c(x) = x$, which meets the inventory storage cost function assumptions made in the definition of (ROP). We will bound $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{S}}(1;\tilde{p}_1,\dots,\tilde{p}_N)\big]$ and $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{M}}(1;\tilde{p}_1,\dots,\tilde{p}_N)\big]$ in the limit as $\beta \to 1$. First, consider $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{S}}(1;\tilde{p}_1,\dots,\tilde{p}_N)\big]$. With $\boldsymbol{V}_{\mathrm{S}}(0;p_1,\dots,p_N)=0$, using the expression for the derivative given by Lemma 4, we write $\boldsymbol{V}_{\mathrm{S}}(1;p_1,\dots,p_N)=\int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{V}_{\mathrm{S}}(x;p_1,\dots,p_N)\mathrm{d}x$. Hence,

$$\mathbb{E}\big[\boldsymbol{V}_{\mathrm{S}}(1;\tilde{p}_1,\dots,\tilde{p}_N)\big] = \mathbb{E}\left[\int_0^1 \frac{\int_{p_{\mathrm{S}}^*(x)}^{\infty} pf(p)\mathrm{d}p - xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x\right]. \tag{35}$$

The negative term in (35) is

$$\mathbb{E}\left[\int_0^1 \frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x\right] = \int_0^{\infty}\cdots\int_0^{\infty}\int_0^1 \frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x f(p_N)\mathrm{d}p_N\cdots f(p_1)\mathrm{d}p_1, \tag{36}$$

where the equality follows from Tonelli's Theorem. Fix $p_1,\dots,p_{N-1}$, and consider the inner-most integral in (36) when p_N is large. The value $p_{\mathrm{S}}^*(x)$ is the p which solves $y_{\mathrm{S}}(\infty,p)=x$. For $\alpha \in (0,1)$, using the expression (10), if p_N is sufficiently large then for each $\beta \in [\alpha,1)$ and $x \in [0,1]$ the value $p_{\mathrm{S}}^*(x)$ is the p which solves $\beta\frac{1}{N}(p_N-p)_+ + (1-\beta)p = x$. It can then be seen that

$$\frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))} \to \frac{-x\left(1-e^{-(p_N-Nx)}\right)}{1-\left(1-e^{-(p_N-Nx)}\right)}$$

uniformly in $x\in[0,1]$ as $\beta\to 1$, whereby

$$\lim_{\beta\to 1}\int_0^1 \frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x = \int_0^1 \frac{-x\left(1-e^{-(p_N-Nx)}\right)}{1-\left(1-e^{-(p_N-Nx)}\right)}\mathrm{d}x = \frac{1}{2}+\left(\frac{1}{N}+\frac{1}{N^2}\right)e^{p_N-N} - \frac{1}{N^2}e^{p_N}.$$

So for each p_N sufficiently large, given $\varepsilon > 0$, there exists a $\beta < 1$ beyond which

$$\int_0^1 \frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x < \frac{1}{2}+\left(\frac{1}{N}+\frac{1}{N^2}\right)e^{p_N-N} - \frac{1}{N^2}e^{p_N}+\varepsilon. \tag{37}$$

But, $f(p_N)=e^{-p_N}$ and

$$\lim_{p_N\to\infty}\left(\frac{1}{2}+\left(\frac{1}{N}+\frac{1}{N^2}\right)e^{p_N-N} - \frac{1}{N^2}e^{p_N}+\varepsilon\right)e^{-p_N} = \left(\frac{1}{N}+\frac{1}{N^2}\right)e^{-N} - \frac{1}{N^2} < 0$$

which means that for $0<a<b$ the integral

$$\int_a^b \left(\frac{1}{2}+\left(\frac{1}{N}+\frac{1}{N^2}\right)e^{p_N-N} - \frac{1}{N^2}e^{p_N}+\varepsilon\right)e^{-p_N}\mathrm{d}p_N$$

can be made arbitrarily negative by increasing b. Using the upper bound (37), it follows that the integral

$$\int_0^{\infty}\int_0^1 \frac{-xF(p_{\mathrm{S}}^*(x))}{1-\beta F(p_{\mathrm{S}}^*(x))}\mathrm{d}x f(p_N)\mathrm{d}p_N$$

in (36) can be made arbitrarily negative by increasing β towards 1, regardless of the values of $p_1,\dots,p_{N-1}$. Hence, (36) is unbounded below.

Given samples $p_1,\ldots,p_N$, recall that $p^*_{\mathrm{S}}(x)\le\max\{p_1,\ldots,p_N\}$. The positive term in (35) is

$$\begin{aligned}\mathbb{E}\left[\int_0^1\frac{\int_{p^*_{\mathrm{S}}(x)}^{\infty}pf(p)\mathrm{d}p}{1-\beta F(p^*_{\mathrm{S}}(x))}\mathrm{d}x\right]&\le\mathbb{E}\left[\int_0^1\frac{\int_{p^*_{\mathrm{S}}(x)}^{\infty}pf(p)\mathrm{d}p}{1-F(p^*_{\mathrm{S}}(x))}\mathrm{d}x\right]\\&=\mathbb{E}\left[\int_0^1 1+p^*_{\mathrm{S}}(x)\mathrm{d}x\right]\\&\le\mathbb{E}\left[\int_0^1 1+\max\{\tilde{p}_1,\ldots,\tilde{p}_N\}\mathrm{d}x\right]\\&=\mathbb{E}\left[1+\max\{\tilde{p}_1,\ldots,\tilde{p}_N\}\right].\end{aligned}\tag{38}$$

The value $\mathbb{E}\big[\max\{\tilde{p}_1,\ldots,\tilde{p}_N\}\big]$ is just the expected value of the N^{th} order statistic of N random samples from P. With P being an Exponential(1) distribution, this is finite (see e.g. David and Nagaraja 2004, Section 4.2). Hence, (38) is bounded above. We have already shown that (36) is unbounded below as $\beta\to 1$, and we have just shown that (38) is bounded above. Thus, their sum (35) is unbounded below in the same limit, regardless of the choice of N.

Consider now $\mathbb{E}\big[\boldsymbol{V}_{\mathrm{M}}(1;\tilde{p}_1,\ldots,\tilde{p}_N)\big]$. For simplicity, we set $N=2$, although the result can be shown to hold for each $N\ge 2$. Following similar reasoning as in (35), we have

$$\mathbb{E}\big[\boldsymbol{V}_{\mathrm{M}}(1;\tilde{p}_1,\tilde{p}_2)\big]=\mathbb{E}\left[\int_0^1\frac{\int_{p^*_{\mathrm{M}}(x)}^{\infty}pf(p)\mathrm{d}p-xF(p^*_{\mathrm{M}}(x))}{1-\beta F(p^*_{\mathrm{M}}(x))}\mathrm{d}x\right]\tag{39}$$

which has negative term

$$\mathbb{E}\left[\int_0^1\frac{-xF(p^*_{\mathrm{M}}(x))}{1-\beta F(p^*_{\mathrm{M}}(x))}\mathrm{d}x\right]=\int_0^\infty\int_0^\infty\int_0^1\frac{-xF(p^*_{\mathrm{M}}(x))}{1-\beta F(p^*_{\mathrm{M}}(x))}\mathrm{d}xf(p_2)\mathrm{d}p_2f(p_1)\mathrm{d}p_1.\tag{40}$$

The iterated integral in (40) can be divided into ranges based on the value of the sample average $\mu_2:=\frac{1}{2}(p_1+p_2)$. Using the expression (11), observe that, depending on μ_2, either $p^*_{\mathrm{M}}(x)=\beta\mu_2-x$ or $p^*_{\mathrm{M}}(x)=-x/(1-\beta)$. Since $\mu_2-x\ge\beta\mu_2-x\ge -x/(1-\beta)$, in (40) the term

$$\begin{aligned}\int_0^1\frac{-xF(p^*_{\mathrm{M}}(x))}{1-\beta F(p^*_{\mathrm{M}}(x))}\mathrm{d}x&\ge\int_0^1\frac{-xF(\mu_2-x)}{1-F(\mu_2-x)}\mathrm{d}x\\&=\int_0^{\min\{\mu_2,1\}}\frac{-x\left(1-e^{-(\mu_2-x)}\right)}{1-\left(1-e^{-(\mu_2-x)}\right)}\mathrm{d}x\\&=\frac{1}{2}(\min\{\mu_2,1\})^2+(1+\min\{\mu_2,1\})e^{\mu_2-\min\{\mu_2,1\}}-e^{\mu_2}\\&\ge -e^{\mu_2}\end{aligned}$$

and consequently (40) is bounded below by

$$\int_0^\infty\int_0^\infty -e^{\frac{1}{2}(p_1+p_2)}e^{-p_2}\mathrm{d}p_2e^{-p_1}\mathrm{d}p_1=-4.$$

Moreover, similar reasoning as for (38) shows that the positive term in (39) is bounded above. Thus, their sum (39) is bounded as $\beta\to 1$. □

Acknowledgments

We would like to acknowledge helpful discussions with Andrew Mason.

References

Anderson EJ, Philpott AB (2022) Improving sample average approximation using distributional robustness. *INFORMS Journal on Optimization* 4(1):90–124.

Bellman R (1957) *Dynamic Programming* (Princeton, NJ: Princeton University Press).

Bertsekas DP (2005) Dynamic programming and suboptimal control: A survey from ADP to MPC. *European Journal of Control* 11(4):310–334.

Bertsekas DP (2022) Newton's method for reinforcement learning and model predictive control. *Results in Control and Optimization* 7:100121.

Bertsekas DP, Shreve SE (1996) *Stochastic Optimal Control: The Discrete-Time Case* (Belmont, MA: Athena Scientific).

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM Review* 59(1):65–98.

David HA, Nagaraja HN (2004) *Order Statistics* (Newark: Wiley InterScience), 3 edition.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

Downward A, Dowson O, Baucke R (2020) Stochastic dual dynamic programming with stagewise-dependent objective uncertainty. *Operations Research Letters* 48:33–39.

Dupačová J (1966) On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky* 91(4):423–430.

Dyer M, Stougie L (2006) Computational complexity of stochastic programming problems. *Mathematical Programming* 106:423–432.

Gotoh J, Kim MJ, Lim AEB (2023) Technical note—a data-driven approach to beating saa out of sample. *Operations Research* Forthcoming.

Iyengar GN (2005) Robust dynamic programming. *Mathematics of Operations Research* 30(2):257–280.

Martin T (2021) *Stochastic optimization for the procurement of crude oil in refineries*. PhD thesis, École des Ponts ParisTech.

Pacaud F, Carpentier P, Chancelier JP, De Lara M (2024) Optimization of a domestic microgrid equipped with solar panel and battery: Model predictive control and stochastic dual dynamic programming approaches. *Energy Systems* 15:115–139.

Pereira M, Pinto L (1991) Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming* 52:359–375.

Scarf HE (1958) A min-max solution of an inventory problem. Arrow KJ, Karlin S, Scarf HE, eds., *Studies in the Mathematical Theory of Inventory and Production*, 201–209 (Redwood City, CA: Stanford University Press).

Shapiro A (2006) On complexity of multistage stochastic programs. *Operations Research Letters* 34(1):1–8.

Shapiro A (2016) Rectangular sets of probability measures. *Operations Research* 64(2):528–541.

Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (Philadelphia, PA: Society for Industrial and Applied Mathematics).

Shapiro A, Ding L (2020) Periodical multistage stochastic programs. *SIAM Journal on Optimization* 30(3):2083–2102.

Shapiro A, Nemirovski A (2005) On complexity of stochastic programming problems. Jeyakumar V, Rubinov A, eds., *Continuous Optimization: Current Trends and Modern Applications*, 111–146 (Boston, MA: Springer).

Stokey NL, Lucas RE Jr, Prescott EC (1989) *Recursive Methods in Economic Dynamics* (Cambridge, MA: Harvard University Press).

Theil H (1957) A note on certainty equivalence in dynamic planning. *Econometrica* 25(2):346–349.

Ziemba WT (1971) Transforming stochastic dynamic programming problems into nonlinear programs. *Management Science* 17(7):450–462.