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 optimization problem that is representative of the type for which MPC may be the better choice. We find that this can indeed be the case when the probability distribution of the underlying random variable is skewed or has enough weight in the right-hand tail.

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 optimization problem to deterministic forecasts, giving a cost-to-go function that is computationally inexpensive to evaluate (Bertsekas 2005). In this way MPC can handle 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 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, but they provide an intuition for the types of problems for which one method may be a better choice than 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 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 as inventory for later sales at a different

realisation of the random 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 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 version. 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 random 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 give some notation. An element $x \in \mathcal{X} \subset \mathbb{R}^n$ denotes the state vector, and an element $\xi \in \Xi \subset \mathbb{R}^m$ denotes the outcome of a random noise vector $\boldsymbol{\xi}$ distributed according to $\mathbb{P}$. (Boldface distinguishes random vectors from their outcomes.) Superscripts denote indexing by stage, but also, where clear from context, raising to a power.

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

$$\underset{y^1, y^2, \dots}{\text{minimize}} \quad \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x^t, x^{t+1}, \boldsymbol{\xi}^t) \right] \tag{SOC}$$

$$\text{subject to} \quad x^{t+1} = y^t(x^t, \boldsymbol{\xi}^t) \in \mathcal{Y}(x^t, \boldsymbol{\xi}^t) \quad \forall t \in \mathbb{N},$$

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

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 next state, 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 mapping $\mathcal{Y} : \mathcal{X} \times \Xi \rightrightarrows \mathcal{X}$ is nonempty- and compact-valued, and continuous.*

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

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

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

For simplicity we do not vary $\mathbb{P}$, φ, or $\mathcal{Y}$ between stages. However, our results can be generalised to such a setting using a periodic formulation of (SOC), as in (Shapiro and Ding 2020). Also, 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 $v : \mathcal{X} \times \Xi \to \mathbb{R}$ which satisfies the functional equation

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

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

$$y(x, \xi) \in \underset{y \in \mathcal{Y}(x,\xi)}{\operatorname{argmin}} \Big\{ \varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{\xi})\big] \Big\} \quad \forall (x, \xi) \in \mathcal{X} \times \Xi.$$

This defines a decision rule for each stage 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 $y : \mathcal{X} \times \Xi \to \mathcal{X}$ with $\xi \mapsto y(x, \xi)$ measurable and $y(x, \xi) \in \mathcal{Y}(x, \xi)$ for every $(x, \xi) \in \mathcal{X} \times \Xi$ as an *admissible policy* for (SOC).

2.1. Data-Driven Stochastic Optimal Control

To approximate (SOC) in a data-driven manner, given N samples $\xi_1, \ldots, \xi_N \in \Xi$ we form the empirical distribution $\mathbb{P}_N$. The infinite-product distribution $\mathbb{P}_N^\infty = \mathbb{P}_N \times \mathbb{P}_N \times \cdots$ then defines an approximation of the joint distribution for the random vectors $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots$, which yields the problem

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

As for (1), this is related to the problem of finding a function $v_{\mathrm{S}} : \mathcal{X} \times \Xi \to \mathbb{R}$ which satisfies the sample-based SDP functional equation

$$v_{\mathrm{S}}(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}_N}\big[v_{\mathrm{S}}(y, \boldsymbol{\xi})\big] \Big\} \quad \forall (x, \xi) \in \mathcal{X} \times \Xi. \tag{3}$$

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 SDP policy $y_{\mathrm{S}} : \mathcal{X} \times \Xi \to \mathcal{X}$ satisfying

$$y_{\mathrm{S}}(x, \xi) \in \underset{y \in \mathcal{Y}(x,\xi)}{\operatorname{argmin}} \Big\{ \varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}_N}\big[v_{\mathrm{S}}(y, \boldsymbol{\xi})\big] \Big\} \quad \forall (x, \xi) \in \mathcal{X} \times \Xi$$

can be obtained.

MPC proceeds differently and replaces the random vectors in (SOC) with a deterministic forecast. Given N samples $\xi_1, \ldots, \xi_N \in \Xi$, we use the sample average $\mu_N := \frac{1}{N} \sum_{i=1}^{N} \xi_i \in \Xi$ as the forecast. This leads to the problem of finding a function $v_{\mathrm{M}} : \mathcal{X} \times \Xi \to \mathbb{R}$ which satisfies the sample-based MPC functional equation

$$v_{\mathrm{M}}(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta v_{\mathrm{M}}(y, \mu_N) \Big\} \quad \forall (x, \xi) \in \mathcal{X} \times \Xi. \tag{4}$$

Observe that (4) is equivalent to (1) with the distribution $\mathbb{1}_{\mu_N}$ used instead of $\mathbb{P}$. (Here $\mathbb{1}_\xi$ denotes the point-mass 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 (4), and by solving (4) a sample-based MPC policy $y_{\mathrm{M}} : \mathcal{X} \times \Xi \to \mathcal{X}$ satisfying

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

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 (3) depends critically on the values of the N samples used. Instead, by hedging against the worst-case distributions in an *ambiguity set* $\mathcal{P}$ informed by the N samples, *distributionally robust optimization* (DRO) provides a means 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 $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots$, which yields the distributionally robust problem

$$\begin{aligned} \underset{y^1,y^2,\ldots}{\operatorname{minimize}} \quad & \underset{\mathbb{Q} \in \mathcal{P}^{\infty}}{\sup} \; \mathbb{E}_{\mathbb{Q}} \left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x^t, x^{t+1}, \boldsymbol{\xi}^t) \right] \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \boldsymbol{\xi}^t) \in \mathcal{Y}(x^t, \boldsymbol{\xi}^t) \quad \forall t \in \mathbb{N}. \end{aligned} \tag{5}$$

The choice of $\mathcal{P}^{\infty}$ in (5) 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 (5) can be solved by finding a function $v_{\mathrm{R}} : \mathcal{X} \times \Xi \to \mathbb{R}$ which satisfies the DRO functional equation

$$v_{\mathrm{R}}(x,\xi) = \underset{y \in \mathcal{Y}(x,\xi)}{\inf} \Big\{ \varphi(x,y,\xi) + \beta \underset{\mathbb{Q} \in \mathcal{P}}{\sup} \mathbb{E}_{\mathbb{Q}} \big[v_{\mathrm{R}}(y,\boldsymbol{\xi}) \big] \Big\} \quad \forall (x,\xi) \in \mathcal{X} \times \Xi. \tag{6}$$

Note that the ambiguity set in (6) is $\mathcal{P}$ so the supremum is over a set of distributions for $\boldsymbol{\xi}$, rather than a set of joint distributions for $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots$ as in (5).

To demonstrate a connection between DRO and MPC, we now analyse the functional operator associated with (4). Let $\mathcal{C}(\mathcal{X} \times \Xi)$ denote the set of bounded and continuous real-valued functions on $\mathcal{X} \times \Xi$. This is a Banach space under the sup norm $\| \cdot \|_{\infty}$. Under Assumption 1, the equation (4) features an MPC Bellman operator $B_{\mathrm{M}} : \mathcal{C}(\mathcal{X} \times \Xi) \to \mathcal{C}(\mathcal{X} \times \Xi)$ which for $f \in \mathcal{C}(\mathcal{X} \times \Xi)$ at $(x,\xi) \in \mathcal{X} \times \Xi$ has value

$$B_{\mathrm{M}}(f)(x,\xi) := \underset{y \in \mathcal{Y}(x,\xi)}{\inf} \Big\{ \varphi(x,y,\xi) + \beta f(y,\mu_N) \Big\}.$$

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 $\mathcal{C}(\mathcal{X} \times \Xi)$ to 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 (4). Further, 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 mapping $\xi \mapsto f(x, \xi)$ is concave (convex) for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$, and which maps to itself under B_{M}.

As we will show, problems with concavity-preserving MPC Bellman operators give rise to sample-based MPC policies that satisfy a DRO interpretation. A sufficient condition for this concavity preservation is provided by the following result.

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

In some circumstances we can establish this key property of concavity preservation without requiring all the conditions of Proposition 1. Some special cases will be discussed in examples below.

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 $\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mu_N$ for every $\mathbb{Q} \in \mathcal{P}$ and that $\mathbb{1}_{\mu_N} \in \mathcal{P}$. If the MPC Bellman operator B_{M} is concavity preserving, then the DRO functional equation (6) has a solution which is the same as the solution to the MPC functional equation (4).*

In contrast to DRO, *distributionally optimistic optimization* (DOO) considers only the best-case distributions in an ambiguity set $\mathcal{P}$. Using the rectangular ambiguity set $\mathcal{P}^{\infty}$, a distributionally optimistic version of (5) can be formed. This leads to the problem of finding a function $v_{\mathrm{O}} : \mathcal{X} \times \Xi \to \mathbb{R}$ which satisfies the DOO functional equation

$$v_{\mathrm{O}}(x, \xi) = \inf_{y \in \mathcal{Y}(x, \xi)} \left\{ \varphi(x, y, \xi) + \beta \inf_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}} \big[v_{\mathrm{O}}(y, \boldsymbol{\xi}) \big] \right\} \quad \forall (x, \xi) \in \mathcal{X} \times \Xi. \tag{7}$$

The essential characteristic of the connection between DRO and MPC is the concavity-preserving nature of the operator B_{M}. If B_{M} is instead 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 rather than concavity preserving. Then the DOO functional equation (7) has a solution which is the same as the solution to the MPC functional equation (4).*

We now give some examples of problems with concavity- and convexity-preserving MPC Bellman operators that differ from those characterised by Proposition 1.

EXAMPLE 1. Downward et al. (2020) study a class of problems where stagewise-dependent uncertainty appears in the objective function (see also Dowson 2018, Chapter 4). These include resource allocation problems under price uncertainty, such as hydroelectric bidding and agricultural production planning. We present an example from this class that minimizes the production cost of a commodity sold on a spot market. Suppose at time t the spot price p^t is an autoregressive process defined by $\boldsymbol{p}^t = \alpha p^{t-1} + \boldsymbol{\xi}$. Further, suppose the commodity has per-period production w and marginal per-period storage cost c. Here the state is determined by the inventory $x \in \mathcal{X}$, and now also by the previous spot price $q \in \mathcal{Q}$. As before the decision variables are the next state: an inventory-price pair (y, p), but in this case p is the current price, already determined by q and ξ. The functional equation is

$$v(x, q, \xi) = \inf_{(y,p) \in \mathcal{Y}(x,q,\xi)} \left\{ -p(x + w - y) + cy + \beta \mathbb{E}\big[v(y, p, \boldsymbol{\xi})\big] \right\} \quad \forall (x, q, \xi) \in \mathcal{X} \times \mathcal{Q} \times \Xi$$

where

$$\mathcal{Y}(x, q, \xi) = \big\{ (y, p) \in \mathcal{X} \times \mathcal{Q} : y \le x + w, \ p = \alpha q + \xi \big\}.$$

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

$$\mathcal{F} = \big\{ f \in \mathcal{C}(\mathcal{X} \times \mathcal{Q} \times \Xi) : (q, \xi) \mapsto f(x, q, \xi) \text{ is concave for every } x \in \mathcal{X} \big\}.$$

For every $f \in \mathcal{F}$ the mapping $p \mapsto -p(x + w - y) + cy + \beta f(y, p, \mathbb{E}[\boldsymbol{\xi}])$ is concave. Due to the equality constraint $p = \alpha q + \xi$, it can be shown that the infimum operation preserves concavity in (q, ξ) through a similar argument to that in the proof of Proposition 1. Thus, sample-based MPC satisfies the DRO 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 $t \in \mathcal{T}$, constant per-period demand d, and a hydro plant that generates $h \in \mathcal{H}$ from a reservoir with storage $x \in \mathcal{X}$ and random inflow $\boldsymbol{\xi}$. All quantities are measured in terms of energy. The functional equation is

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

where

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

Here h and t are viewed as dummy state variables in our framework. To see that the MPC Bellman operator is convexity preserving, consider the set

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

For every $f \in \mathcal{F}$ the mapping $(y,h,t) \mapsto t + \beta f(y, \mathbb{E}[\boldsymbol{\xi}])$ is convex. Due to the linearity of the constraints, it can be shown that the infimum operation preserves convexity in (x,ξ) through a similar argument to that in the proof of Proposition 1 (although with the inequalities reversed). Thus, sample-based MPC satisfies the DOO interpretation of Corollary 1. $\quad\square$

If the MPC Bellman operator is concavity preserving, Theorem 1 shows that sample-based MPC can be interpreted as a distributionally robust version of sample-based SDP. Distributional robustness can sometimes 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 sample-based MPC can be interpreted as a distributionally optimistic version of sample-based SDP. Distributional optimism can sometimes worsen out-of-sample performance due to increased solution variance (Gotoh et al. 2023). This suggests that for small sample sizes, when the MPC Bellman operator is concavity preserving as in Example 1, MPC may be a better choice than SDP. On the other hand, when the MPC Bellman operator is convexity preserving as in Example 2, SDP may be a better choice than MPC. In the following section we provide performance guarantees which support these observations.

4. Out-of-Sample Performance

Suppose that the distribution $\mathbb{P}$ has a finite mean $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$. For N independent random samples $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N \sim \mathbb{P}$, the sample average $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\xi}_i$ is usually a good estimate of the true mean $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$, as long as N is sufficiently large. Indeed, the strong law of large numbers shows $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\xi}_i \to \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ almost surely as $N \to \infty$. MPC can take advantage of this situation since the computational complexity of solving (4) does not depend on the sample size, while SDP cannot.

We now show that, when the true mean is known and used as the MPC forecast, the solution to the MPC functional equation (4) 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 $\mu_N = \mu := \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ in the MPC functional equation (4). If the resulting MPC Bellman operator B_{M} is concavity preserving, then*

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

That is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (4) is an upper bound on the cost incurred when applying the resulting policy to (SOC).

If the sample size is large then using MPC with the sample mean as the forecast is likely to give a result close to that with the true mean. This suggests that applying the sample-based MPC policy to (SOC) is unlikely to cause 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, so applying MPC with the true mean used as the forecast will do worse than predicted.

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*

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

That is, the value $v_{\mathrm{M}}(x^1, \mu)$ obtained by solving the MPC functional equation (4) is a lower bound on the cost incurred when applying the resulting policy to (SOC).

Corollary 2 suggests that sample-based MPC is likely to cause out-of-sample disappointment when the MPC Bellman operator is convexity preserving.

In the rest of this section we derive a result for comparing the cost of applying different policies to (SOC). Given an admissible policy $y : \mathcal{X} \times \Xi \to \mathcal{X}$, with φ continuous, the mapping $\xi \mapsto \varphi(x, y(x, \xi), \xi)$ is measurable for every $x \in \mathcal{X}$. Hence, the objective value of (SOC) under y is well defined. Let $\widetilde{V}_y : \mathcal{X} \to \mathbb{R}$ denote this value as a function of the initial state, so

$$\widetilde{V}_y(x^1) := \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^\infty \beta^{t-1}\varphi(x^t, x^{t+1}, \boldsymbol{\xi}^t)\right] \quad \text{where } x^{t+1} = y(x^t, \boldsymbol{\xi}^t)\, \forall t \in \mathbb{N}. \tag{8}$$

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

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

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

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

Under Assumption 2(iii), the tth term in the sum in (8) is bounded by the random variable $l(\boldsymbol{\xi}^t)$ which has a finite expectation. With $\beta \in (0, 1)$, it follows that $\widetilde{V}_y$ is bounded. Having defined $\widetilde{V}_y$ as a function of the initial state and shown that it is bounded, it satisfies the functional equation

$$\widetilde{V}_y(x) = \mathbb{E}_{\mathbb{P}}\left[\varphi(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta\widetilde{V}_y(y(x, \boldsymbol{\xi}))\right] \quad \forall x \in \mathcal{X}. \tag{9}$$

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. For two admissible policies $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, define $\widetilde{V}_{y,y'} : \mathcal{X} \to \mathbb{R}$ by

$$\widetilde{V}_{y,y'}(x) := \mathbb{E}_{\mathbb{P}}\Big[\varphi\big(x, y(x,\boldsymbol{\xi}), \boldsymbol{\xi}\big) + \beta \widetilde{V}_{y'}\big(y(x,\boldsymbol{\xi})\big)\Big].$$

The value $\widetilde{V}_{y,y'}(x)$ is the objective value of (SOC) when starting from initial state x if policy y is followed in the first stage and policy y' is followed thereafter. It is clear that $\widetilde{V}_{y,y'}$ is well defined and bounded in the same way that $\widetilde{V}_y$ and $\widetilde{V}_{y'}$ are.

THEOREM 3. *Let Assumption 2 hold. For two admissible policies $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, if*

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

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

Rather than having to calculate integrals directly to compare the performance of different policies, Theorem 3 allows us to check a uniform condition involving similarly defined functional equations.

5. Revenue Optimization With Stochastic Prices

To gain a deeper understanding of the differences between sample-based SDP and MPC, we study a particular problem from within the setting considered in Section 2. This problem is an example from the class for which sample-based MPC satisfies the DRO interpretation of Theorem 1. For initial inventory level $x^1 \in \mathbb{R}_+$ we consider the revenue optimization problem

$$\underset{y^1,y^2,\dots}{\text{maximize}} \quad \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1}\big(\boldsymbol{p}^t(x^t - x^{t+1}) - C(x^{t+1})\big)\right] \qquad \text{(ROSP)}$$

$$\text{subject to} \quad x^{t+1} = y^t(x^t, \boldsymbol{p}^t) \in [0, x^t] \quad \forall t \in \mathbb{N}.$$

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 realisation of the random variable $\boldsymbol{p}^t$. Here $\boldsymbol{p}^1, \boldsymbol{p}^2, \dots$ are random prices that are independent and identically distributed according to $\mathbb{P}$, and $C : \mathbb{R}_+ \to \mathbb{R}_+$ is an inventory-storage cost function. We assume that C is increasing, strictly convex, and continuously differentiable. Writing $c(x) := \frac{d}{dx}C(x)$, with c strictly increasing and continuous, we further define a continuous inverse function c^{-1} on the range of c. Since the inventory levels are restricted to the compact set $[0, x^1]$, without loss of generality we also assume that $\lim_{x\to\infty} c(x) = \infty$.

The problem (ROSP) is 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 realisation of the random price $\boldsymbol{p}^t$, while incurring a storage cost $C(x^{t+1})$ on their remaining inventory. This can model a number of situations. For example, it could describe an electricity distributor with a charged battery deciding when to dispatch electricity, or an investor deciding when to sell a holding of shares.

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

$$v(x,p) = \sup_{0 \le y \le x} \Big\{ p(x-y) - C(y) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{p}) \big] \Big\} \quad \forall (x,p) \in \mathbb{R}_+ \times \mathbb{R}. \tag{10}$$

The constraints in (ROSP) restrict states to the compact set $[0, x^1]$, and with $\mathbb{P}$ having compact support, the terms in the sum in the objective function are bounded. It follows that (10) has a unique continuous solution $v : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and that the mapping $x \mapsto v(x,p)$ is concave for every $p \in \mathbb{R}$ (Stokey et al. 1989, Chapter 9). Given such v, a policy $y : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ satisfying

$$y(x,p) \in \operatorname*{argmax}_{0 \le y \le x} \Big\{ p(x-y) - C(y) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{p}) \big] \Big\} \quad \forall (x,p) \in \mathbb{R}_+ \times \mathbb{R} \tag{11}$$

is optimal for (ROSP).

In the notation of Section 2, (ROSP) has $\mathcal{Y}(x,p) = [0,x]$, and a suitable change of perspective from maximization to minimization yields the cost function $\varphi(x,y,p) = -p(x-y) + C(y)$ which is concave in p. This satisfies the conditions of Proposition 1, so applying sample-based MPC to (ROSP) will yield the DRO interpretation of Theorem 1.

Denote the projection of $z \in \mathbb{R}$ onto $[a,b] \subset \mathbb{R}$ by $(z)_{[a,b]}$ and write $(z)_+$ for its projection onto the nonnegative real numbers. Using the inclusion (11) we arrive at the following proposition which presents the optimal policy in closed form.

PROPOSITION 2. *Suppose that the distribution $\mathbb{P}$ has compact support. Then the policy which at inventory x and price p sells down to*

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

is optimal for (ROSP).

The optimal policy has a natural interpretation: the term $\beta \mathbb{E}_{\mathbb{P}}\big[(\boldsymbol{p} - p)_+ \big]$ 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 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 storage cost.

Proposition 2 shows that for each price p the problem (ROSP) has an optimal target inventory level $y(\infty, p) := \lim_{x \to \infty} y(x,p)$. At inventory 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 minimum acceptable price $p^\star(x)$ 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 c^{-1}, the function $p^\star$ is also continuous.

Proposition 2 makes no assumptions about the distribution $\mathbb{P}$, except that it has compact support. Thus, it could have a density on a compact set, or could be an empirical distribution. Using the empirical distribution on N price samples $p_1, \ldots, p_N$, Proposition 2 shows that the optimal SDP policy, which we denote by y_S, is

$$y_\mathrm{S}(x, p; p_1, \ldots, p_N) = c^{-1}\Big(\big(\beta \tfrac{1}{N}\textstyle\sum_{i=1}^N (p_i - p)_+ - (1 - \beta)p\big)_{[c(0), c(x)]}\Big). \tag{12}$$

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

$$y_\mathrm{M}(x, p; p_1, \ldots, p_N) = c^{-1}\Big(\big(\beta(\mu_N - p)_+ - (1 - \beta)p\big)_{[c(0), c(x)]}\Big). \tag{13}$$

In what follows, for simplicity we often suppress the dependence on the price samples $p_1, \ldots, p_N$. We denote by $y_\mathrm{S}(\infty, p)$ and $y_\mathrm{M}(\infty, p)$ the target inventory levels of the policies (12) and (13), and by $p_\mathrm{S}^\star(x)$ and $p_\mathrm{M}^\star(x)$ their minimum acceptable prices required for sales. Clearly it holds that $p_\mathrm{S}^\star \leq \max\{p_1, \ldots, p_N\}$, since this is the highest price the SDP policy considers a possibility. Similar remarks hold for the MPC policy.

Depending on the values of the price samples $p_1, \ldots, p_N$, the sample-based policies may hold on to inventory for too long and overpay for storage, or sell inventory prematurely and not be able to capitalise on higher prices offered in the future. Jensen's Inequality shows that $\sum_{i=1}^N (p_i - p)_+ \geq (\mu_N - p)_+$, and thus $y_\mathrm{S}(\infty, p) \geq y_\mathrm{M}(\infty, p)$. This reveals that the MPC policy decides that it is worthwhile to sell at lower prices than the SDP policy does. Indeed, $p_\mathrm{S}^\star(x) \geq p_\mathrm{M}^\star(x)$ for every inventory level x, and the trade-off between overpaying for storage and selling inventory prematurely is handled differently by sample-based SDP and MPC.

5.1. Out-of-Sample Performance

To study the out-of-sample performance of SDP and MPC on (ROSP), we make the following assumption regarding the true distribution.

ASSUMPTION 3. *The distribution $\mathbb{P}$ has nonnegative support, a finite mean, and no atoms.*

Under Assumption 3, for every $x, y \in [0, x^1]$ and $\mathbb{P}$-almost every $p \in \mathbb{R}_+$, it holds that

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

With $\mathbb{E}_\mathbb{P}\big[\boldsymbol{p}x^1 + C(x^1)\big]$ finite valued, this shows that Assumption 3 for (ROSP) is akin to Assumption 2 for (SOC), which enables us to define the out-of-sample performance of policies using functional equations.

Under Assumption 3 the out-of-sample performance of the SDP policy (12) when starting from initial inventory x, which we denote by $\widetilde{V}_{\mathrm{S}}(x)$, is well defined and satisfies the functional equation

$$\widetilde{V}_{\mathrm{S}}(x) = \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\big(x - y_{\mathrm{S}}(x,\boldsymbol{p})\big) - C\big(y_{\mathrm{S}}(x,\boldsymbol{p})\big) + \beta\widetilde{V}_{\mathrm{S}}\big(y_{\mathrm{S}}(x,\boldsymbol{p})\big)\Big]. \tag{14}$$

Similarly, the out-of-sample performance of the MPC policy (13) when starting from initial inventory x, which we denote by $\widetilde{V}_{\mathrm{M}}(x)$, is well defined and satisfies the functional equation

$$\widetilde{V}_{\mathrm{M}}(x) = \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\big(x - y_{\mathrm{M}}(x,\boldsymbol{p})\big) - C\big(y_{\mathrm{M}}(x,\boldsymbol{p})\big) + \beta\widetilde{V}_{\mathrm{M}}\big(y_{\mathrm{M}}(x,\boldsymbol{p})\big)\Big]. \tag{15}$$

Let $\mathbb{1}$ be the event indicator; that is $\mathbb{1}\{\mathcal{E}\} = 1$ if the event $\mathcal{E}$ is true and $\mathbb{1}\{\mathcal{E}\} = 0$ if it is false. We use Theorem 3 to establish the following proposition which compares the out-of-sample performance of SDP and MPC.

PROPOSITION 3. *Let Assumption 3 hold. For price samples $p_1, \ldots, p_N$ which determine the sample-based SDP and MPC policies by (12) and (13), if the minimum acceptable SDP price $p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)$ is such that*

$$c(x) \geq \beta\mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{\boldsymbol{p} > p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)\big\}\Big] \quad \forall x \in [0, x^1],$$

then $\widetilde{V}_{\mathrm{M}}(x; p_1, \ldots, p_N) \geq \widetilde{V}_{\mathrm{S}}(x; p_1, \ldots, p_N)$ for each $x \in [0, x^1]$. That is, the MPC policy performs at least as well out of sample on (ROSP) as the SDP policy.

The condition has a natural interpretation: $\beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)\}\big]$ represents the expected revenue that SDP gains over MPC by delaying sales, and if this is less than the marginal cost of storage, then waiting provides no net benefit. Without loss of generality, assume $p_1 \leq \cdots \leq p_N$ and recall that $p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N) \leq p_N$. The mapping $p_N \mapsto (p_N - p)_+$ in (12) is strictly increasing at the point $p = p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)$, where $y_{\mathrm{S}}(\infty, p; p_1, \ldots, p_N) = x$. It follows that $p_N \mapsto p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)$ is strictly increasing, and thus $p_N \mapsto \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{S}}^\star(x; p_1, \ldots, p_N)\}\big]$ is decreasing. If the N price samples are independent and identically distributed according to $\mathbb{P}$, when $\mathbb{P}$ has a small amount of probability at very high prices, we will occasionally sample a maximal price that is sufficiently large for the inequality to hold.

Proposition 3 is not an explicit statement about the relative expected out-of-sample performance of SDP and MPC under the sampling distribution. In fact, in most applications it is likely that there will always be some outcomes of the 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 form of true price distribution: the sample-based SDP policy can be misled by overly high price samples that cause it to hold on to inventory for too long and overpay for storage. Consequently, right skew and the size of the right-hand tail in the true distribution

are likely to impact the performance of SDP. On the other hand, due to its distributionally robust properties, the sample-based MPC policy more quickly sells down to lower inventory levels and is protected against this.

If $c(0) = 0$, Proposition 3 does not apply. But this is to be expected; when $c(0) = 0$ an infinitesimal amount of inventory incurs negligible storage costs, so waiting longer under the sample-based SDP policy to observe a higher price will always perform better in the long run. Despite this, for nonnegligible initial inventory levels, we present examples below which show that MPC can still outperform SDP when $c(0) = 0$, due to high storage costs accumulated early on.

5.2. Expected Out-of-Sample Performance

In the rest of this section we present examples which compare the expected out-of-sample performance of SDP and MPC on (ROSP) for price distributions with different skews and tail sizes. We suppose that the N price samples $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N$ used to determine the sample-based policies by (12) and (13) are independent and identically distributed according to $\mathbb{P}$. Expected out-of-sample performance is then given by the expectations of (14) and (15) under the sampling distribution of $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N$.

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

PROPOSITION 4. *Let $\mathbb{P}$ be the* 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 SDP on (ROSP) is unbounded below, while that of MPC is bounded.*

Proposition 4 shows that for each sample size greater than or equal to 2 the expected out-of-sample performance of SDP can be made arbitrarily worse than that of MPC by choosing a discount factor that is sufficiently close to 1. For a particular discount factor the performance of SDP could be improved by increasing the sample size, but this 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 sample sizes. Moreover, in real applications the sample-based SDP policy must be computed numerically, but algorithms for doing this cannot handle large sample sizes. MPC does not have this problem.

To compute the expected out-of-sample performance of SDP and MPC on (ROSP) for different price distributions, we use a simulation coded in `Julia` (Bezanson et al. 2017). Although (ROSP) has an infinite horizon, simulation with a finite number of stages (say T) will provide accurate performance estimates as long as this is sufficiently large; we set $T = 1 \cdot 10^3$.[1] With $x^1 = 1$, $\beta = 0.99$, and $C(x) = \frac{1}{2}x^2$, for a given sample size N we repeat:

---

[1] The resulting truncation error has an order less than $0.99^{1000}/(1 - 0.99) \approx 1 \cdot 10^{-3}$.

0. Sample $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N \sim \mathbb{P}$ to determine a policy y using (12) for SDP and (13) for MPC.

1. Sample $\boldsymbol{p}^t \sim \mathbb{P}$, compute $\beta^{t-1}\big(\boldsymbol{p}^t(x^t - y(x^t, \boldsymbol{p}^t)) - C(y(x^t, \boldsymbol{p}^t))\big)$, and set $x^{t+1} = y(x^t, \boldsymbol{p}^t)$.

2. Repeat 1. from stage $t = 1$ to T and compute $\sum_{t=1}^T \beta^{t-1}\big(\boldsymbol{p}^t(x^t - y(x^t, \boldsymbol{p}^t)) - C(y(x^t, \boldsymbol{p}^t))\big)$.

Each repetition of 0.–2. generates a random out-of-sample performance realisation, and the average of these realisations provides a statistical estimate of expected out-of-sample performance. In the following simulations we used $1 \cdot 10^5$ realisations, finding this sufficient to ensure accurate results. (In Figures 1–5 the standard error ranges are smaller than the markers so are not shown.)

5.3. Skewed Price Distributions
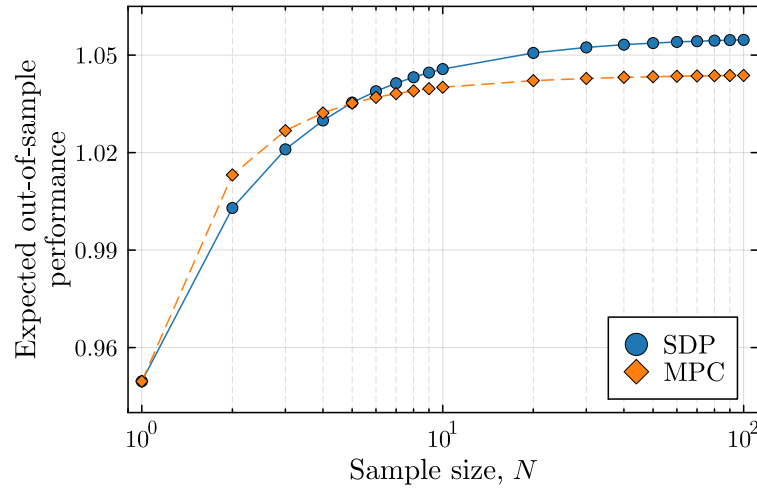
Let Triangular(a, m, b) denote the triangular distribution with lower limit a, mode m, and upper limit b. Figures 1 and 2 present the expected out-of-sample performance of SDP and MPC for left- and right-skewed triangular price distributions with mean 1 and variance $1/8$.
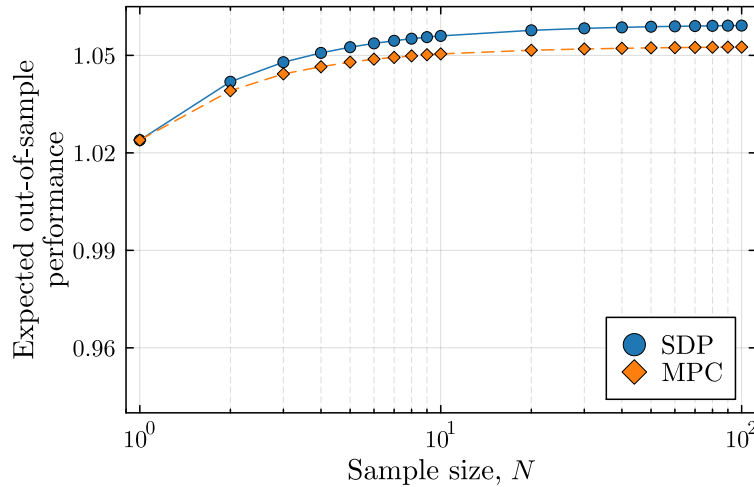
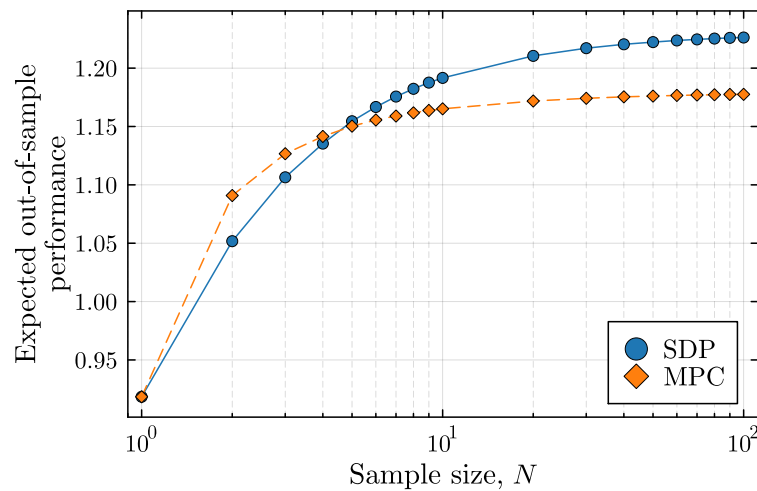**Figure 1**   Performance of SDP and MPC on (ROSP) for Triangular$(0, 3/2, 3/2)$ distributed price.

Figure 2 Performance of SDP and MPC on (ROSP) for Triangular($1/2, 1/2, 2$) distributed price.

Figure 1 shows SDP outperforming MPC for all sample sizes when the true price distribution is left skewed. In contrast, Figure 2 shows MPC outperforming SDP for sample sizes $N \leq 4$ when the true price distribution is right skewed. This reflects the fact that the likelihood of the samples containing a price high enough to cause SDP to undersell inventory and overpay for storage is more likely in the presence of right skew.

Figure 3 presents the expected out-of-sample performance of SDP and MPC for a right-skewed triangular price distribution with mean 1 and variance $1/2$.

Figure 3 Performance of SDP and MPC on (ROSP) for Triangular($0, 0, 3$) distributed price.

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

5.4. Right-Tailed Price Distributions

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

Figure 4   Performance of SDP and MPC on (ROSP) for Exponential(1) distributed price.

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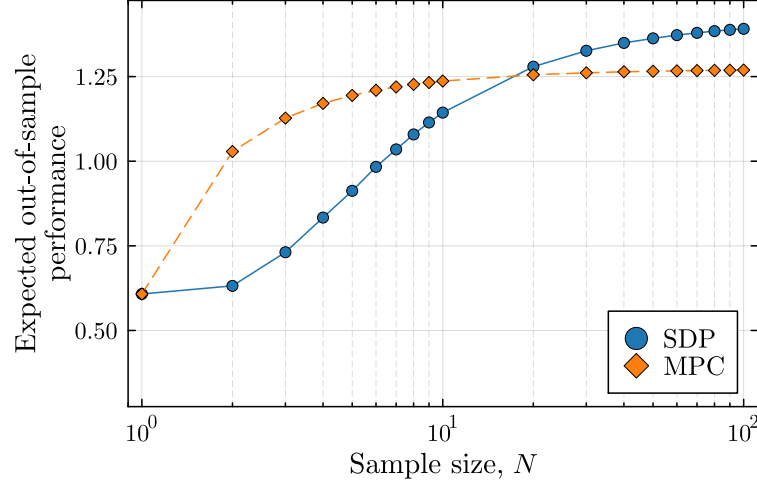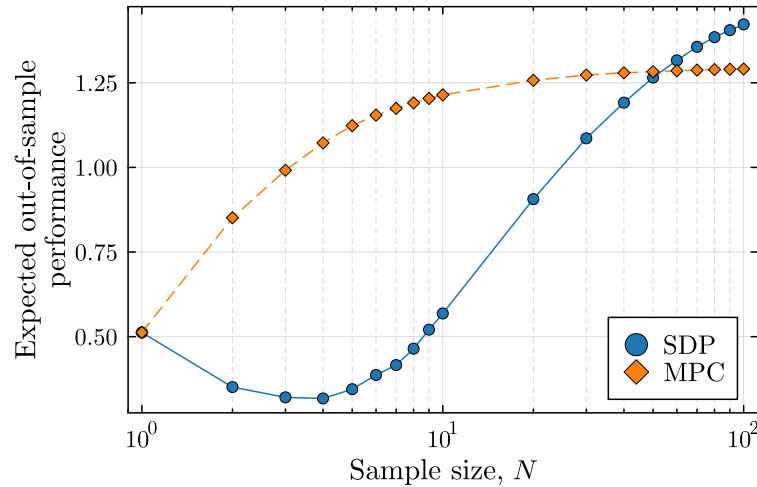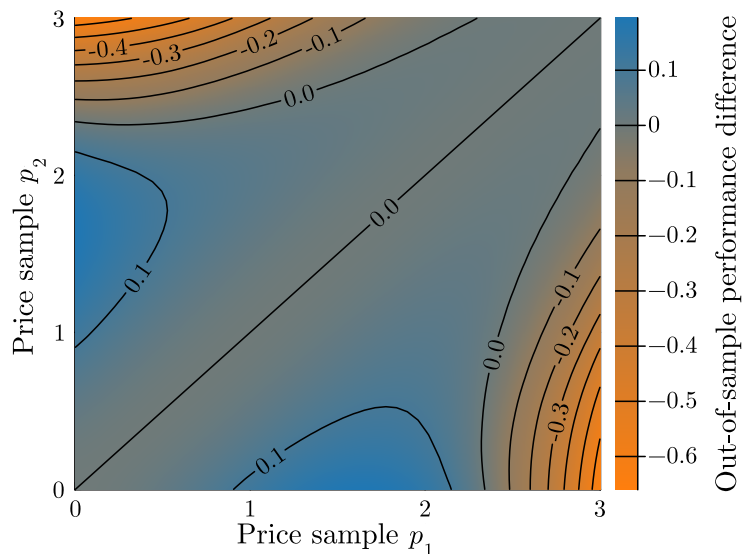

Figure 5   Performance of SDP and MPC on (ROSP) for LogNormal($-1/2, 1$) distributed price.

Figure 5 shows MPC outperforming SDP for samples sizes $N \leq 50$ when the true price distribution is right skewed and has a heavier right-hand tail the exponential distribution used in Figure 4. The poor performance of SDP exhibited here 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 weight in their tails than exponential distributions, this would not be surprising.

For a sample of size $N = 2$, Figure 6 presents the difference in the out-of-sample performance of the SDP and MPC policies.

Figure 6 Difference in SDP and MPC performance on (ROSP) for LogNormal$(-1/2, 1)$ distributed price and $N = 2$ samples. Positive values indicate when SDP outperforms MPC and negative values indicate the reverse.

Figure 6 explicitly demonstrates that a single overly 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 extremes where one sample is very high, the reverse occurs, as predicted by Proposition 3.

6. Discussion

In this paper we provide 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 operator associated with the problem being concavity preserving. When the operator is instead convexity preserving, the interpretation of MPC 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 within 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 study the out-of-sample performance of SDP and MPC on a 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 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 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 such a rolling window includes a very high price sample, the same performance issues will occur.

The stochastic revenue optimization problem we consider 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 an initial level arrives in each time period with probability γ, 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 $\beta = 1 - \gamma$.

Appendix A: Proof of Proposition 1

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

Proof of Proposition 1   Consider the set

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

By definition, $\xi \mapsto f(x, \xi)$ is concave for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$. It remains to show 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 every $x \in \mathcal{X}$, $\xi, \xi' \in \Xi$, and $\lambda \in [0, 1]$, it holds that

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

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

Appendix B: Proof of Theorem 1 and Corollary 1

THEOREM 1. *Let Assumption 1 hold. For an ambiguity set $\mathcal{P}$, suppose that $\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mu_N$ for every $\mathbb{Q} \in \mathcal{P}$ and that $\mathbb{1}_{\mu_N} \in \mathcal{P}$. If the MPC Bellman operator B_{M} is concavity preserving, then the DRO functional equation (6) has a solution which is the same as the solution to the MPC functional equation (4).*

Proof of Theorem 1   The equation (6) 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

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

With B_{M} concavity preserving, there exists a closed subset $\mathcal{F} \subset \mathcal{C}(\mathcal{X} \times \Xi)$ where the mapping $\xi \mapsto f(x, \xi)$ is concave for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$, and which maps to itself under B_{M}. We first show that B_{R} is a contraction mapping on $\mathcal{F}$ via comparison to B_{M}.

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

$$
\mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] \leq f(y, \mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}]) = f(y, \mu_N) \quad \forall \mathbb{Q} \in \mathcal{P}, \tag{16}
$$

where the inequality follows from Jensen's inequality (due to concavity), and the equality follows since $\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mu_N$ for all $\mathbb{Q} \in \mathcal{P}$. The statement (16) implies that $\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] \leq f(y, \mu_N)$. But, with $\mathbb{E}_{\mathbb{1}_{\mu_N}}\big[f(y, \boldsymbol{\xi})\big] = f(y, \mu_N)$ and $\mathbb{1}_{\mu_N} \in \mathcal{P}$, this supremum is attained by $\mathbb{1}_{\mu_N}$ and $\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] = f(y, \mu_N)$. Thus, for every $f \in \mathcal{F}$, we have that

$$
\begin{aligned}
B_{\mathrm{R}}(f)(x, \xi) &= \inf_{y \in \mathcal{Y}(x, \xi)} \Big\{ \varphi(x, y, \xi) + \beta \sup_{Q \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] \Big\} \\
&= \inf_{y \in \mathcal{Y}(x, \xi)} \Big\{ \varphi(x, y, \xi) + \beta f(y, \mu_N) \Big\} = B_{\mathrm{M}}(f)(x, \xi) \quad \forall (x, \xi) \in \mathcal{X} \times \Xi, \tag{17}
\end{aligned}
$$

i.e., B_{R} agrees with B_{M} on $\mathcal{F}$. Recalling that B_{M} is 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 (6) has a solution which is the same as the solution to the MPC functional equation (4). Observe that $(\mathcal{F}, \|\cdot\|_\infty)$ is a Banach space since $\mathcal{F}$ is closed. As 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 (6). In light of (17), this solution is the same as the solution to (4). $\quad\square$

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 (7) has a solution which is the same as the solution to the MPC functional equation (4).*

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

Appendix C: Proof of Theorem 2 and Corollary 2

THEOREM 2. *Let Assumption 1 hold, and set $\mu_N = \mu := \mathbb{E}_\mathbb{P}[\boldsymbol{\xi}]$ in the MPC functional equation (4). If the resulting MPC Bellman operator B_M is concavity preserving, then*

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

That is, the value $v_\mathrm{M}(x^1, \mu)$ obtained by solving the MPC functional equation (4) is an upper bound on the cost incurred when applying the resulting MPC policy to (SOC).

Proof   Let $\mathcal{P} = \{\mathbb{P}, \mathbb{1}_\mu\}$, and recall that $\mathcal{P}^\infty = \mathcal{P} \times \mathcal{P} \times \cdots$ is a rectangular set. By Theorem 1, the MPC policy solves (5), and we have that

$$v_\mathrm{M}(x^1, \mu) = \inf_{y^1, y^2, \cdots} \sup_{\mathbb{Q}\in\mathcal{P}^\infty} \mathbb{E}_\mathbb{Q}\left[\sum_{t=1}^\infty \beta^{t-1}\varphi(x^t, x^{t+1}, \boldsymbol{\xi}^t)\right] \quad \text{subject to } x^{t+1} = y^t(x^t, \boldsymbol{\xi}^t)\,\forall t \in \mathbb{N}$$

$$= \sup_{\mathbb{Q}\in\mathcal{P}^\infty} \mathbb{E}_\mathbb{Q}\left[\sum_{t=1}^\infty \beta^{t-1}\varphi(x^t, x^{t+1}, \boldsymbol{\xi}^t)\right] \quad \text{where } x^{t+1} = y_\mathrm{M}(x^t, \boldsymbol{\xi}^t)\,\forall t \in \mathbb{N}. \tag{18}$$

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

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

which holds since $\mathbb{P}^\infty \in \mathcal{P}^\infty$. The right-hand side equals $v_\mathrm{M}(x^1, \mu)$, by virtue of (18). $\quad\square$

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*

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

that is, the value $v_\mathrm{M}(x^1, \mu)$ obtained by solving the MPC functional equation (4) is a lower bound on the cost incurred when applying the resulting 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. $\quad\square$

Appendix D: Proof of Theorem 3

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

DEFINITION 3. For two admissible policies $y, y' : \mathcal{X} \times \Xi \mapsto \mathcal{X}$, let $\widetilde{V}^1_{y,y'}(x) := \widetilde{V}_{y,y'}(x)$, and for each $T \in \mathbb{N}$ define $\widetilde{V}^{T+1}_{y,y'} : \mathcal{X} \to \mathbb{R}$ by

$$\widetilde{V}^{T+1}_{y,y'}(x) := \mathbb{E}_\mathbb{P}\Big[\varphi\big(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}\big) + \beta \widetilde{V}^T_{y,y'}\big(y(x, \boldsymbol{\xi})\big)\Big].$$

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

LEMMA 1. *Let Assumption 2 hold. For two admissible policies* $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, *it holds that* $\lim_{T \to \infty}\big|\widetilde{V}^T_{y,y'}(x) - \widetilde{V}_y(x)\big| = 0$ *for every* $x \in \mathcal{X}$.

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

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

Continuing this reasoning, an induction shows that

$$\big|\widetilde{V}^T_{y,y'}(x^1) - \widetilde{V}_y(x^1)\big| = \beta^T\Big|\mathbb{E}_{\mathbb{P}^\infty}\big[\widetilde{V}_{y'}(x^T) - \widetilde{V}_y(x^T)\big]\Big| \quad \text{where } x^{t+1} = y(x^t, \boldsymbol{\xi}^t)\, \forall t \in \{1, \ldots, T-1\}. \quad (19)$$

Since $\widetilde{V}_y$ and $\widetilde{V}_{y'}$ are bounded under Assumption 2 and $\beta^T \to 0$ as $T \to \infty$, the value of (19) vanishes as $T \to \infty$. Replacing x^1 with x concludes the proof. $\square$

THEOREM 3. *Let Assumption 2 hold. For two admissible policies* $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, *if*

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

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

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

$$\begin{aligned}
\widetilde{V}^t_{y,y'}(x) &= \mathbb{E}_\mathbb{P}\Big[\varphi\big(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}\big) + \beta \widetilde{V}^{t-1}_{y,y'}\big(y(x, \boldsymbol{\xi})\big)\Big] \\
&\le \mathbb{E}_\mathbb{P}\Big[\varphi\big(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}\big) + \beta \widetilde{V}^t_{y,y'}\big(y(x, \boldsymbol{\xi})\big)\Big] = \widetilde{V}^{t+1}_{y,y'}(x) \quad (20)
\end{aligned}$$

for all $x \in \mathcal{X}$. The inequality (20) establishes the induction. Hence, $\widetilde{V}_{y'}(x) \le \widetilde{V}^t_{y,y'}(x)$ for each $t \in \mathbb{N}$ and $\widetilde{V}_{y'}(x) \le \lim_{t \to \infty} \widetilde{V}^t_{y,y'}(x) = \widetilde{V}_y(x)$, where Lemma 1 yields the final equality. $\square$

Appendix E: Proof of Proposition 2

PROPOSITION 2. *Suppose that the distribution $\mathbb{P}$ has compact support. Then the policy which at inventory x and price p sells down to*

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

is optimal for (ROSP).

Proof of Proposition 2.   Recall that there is a unique continuous function $v : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ satisfying (10) and that $x \mapsto v(x,p)$ is concave for each $p \in \mathbb{R}$. Given such v we derive a policy which satisfies (11) and therefore solves (ROSP). Define $V : \mathbb{R}_+ \to \mathbb{R}$ by $V(x) := \mathbb{E}_{\mathbb{P}}\big[v(x,\boldsymbol{p})\big]$ and note that V is continuous and concave (Shapiro et al. 2009, Theorems 7.43, 7.46). Satisfying (11) amounts to solving

$$\underset{y \in \mathbb{R}_+}{\text{maximize}} \quad p(x-y) - C(y) + \beta V(y) \quad \text{subject to} \quad y \in [0,x]. \tag{21}$$

Due to the strict concavity of C, (21) has a unique solution $y^\star(x,p) \in [0,x]$. Let $y^\star(\infty,p) := \lim_{x\to\infty} y^\star(x,p)$. (This limit is finite valued since $\lim_{y\to\infty} -c(y) = -\infty$.) Now, $y^\star(\infty,p)$ is equal to the optimal solution $y^\star(x,p)$ when projected onto $[0,x]$, and the mapping $p \mapsto y^\star(\infty,p)$ is decreasing. Thus there exists $p^\star(x) \in [-\infty,\infty]$ such that for each $p \in \mathbb{R}$, if $p \leq p^\star(x)$, then $y^\star(\infty,p) \geq x$, and otherwise if $p \geq p^\star(x)$, then $y^\star(\infty,p) \leq x$. Comparing $p^\star(x)$ and p characterises whether or not $y^\star(\infty,p)$ needs to be projected onto $[0,x]$ to solve (21).

We proceed by deriving an expression for the derivative of V using a fixed-point equation. In fact we use the superdifferential operator ∂ since V may be nonsmooth. Due to continuity and concavity, the superdifferential $\partial V(x)$ is nonempty and compact. We also denote the superdifferential of v with respect to x by $\partial_x v(x,p)$. When $p \leq p^\star(x)$, by definition $y^\star(\infty,p) \geq x$, and thus $y^\star(x,p) = x$. In view of (10),

$$v(x,p) = -C(x) + \beta V(x) \quad \forall p \leq p^\star(x)$$

$$\implies -c(x) + \beta \partial V(x) \subset \partial_x v(x,p). \tag{22}$$

On the other hand,

$$v(x,p) = p\big(x - y^\star(\infty,p)\big) - C\big(y^\star(\infty,p)\big) + \beta V\big(y^\star(\infty,p)\big) \quad \forall p \geq p^\star(x)$$

$$\implies p \in \partial_x v(x,p). \tag{23}$$

We therefore define $\psi_x : \partial V(x) \times \mathbb{R} \to \mathbb{R}$ by

$$\psi_x(g,p) := \begin{cases} -c(x) + \beta g & \text{if } p \leq p^\star(x) \\ p & \text{if } p > p^\star(x); \end{cases} \tag{24}$$

the two cases here coinciding with (22) and (23). Hence, for each $g \in \partial V(x)$,

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

and thus $\mathbb{E}_{\mathbb{P}}\big[\psi_x(g,\boldsymbol{p})\big] \in \partial V(x)$ (Shapiro et al. 2009, Theorem 7.47). The expectation of (24) is

$$\mathbb{E}_{\mathbb{P}}\big[\psi_x(g,\boldsymbol{p})\big] = \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \leq p^\star(x)\}\big]\big(\beta g - c(x)\big) + \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}\big], \tag{25}$$

and for $g, g' \in \partial V(x)$ taking the difference shows that

$$\left| \mathbb{E}_{\mathbb{P}}\big[\psi_x(g, \boldsymbol{p})\big] - \mathbb{E}_{\mathbb{P}}\big[\psi_x(g', \boldsymbol{p})\big] \right| = \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p^\star(x)\}\big]\beta|g - g'| < |g - g'|,$$

i.e., $g \mapsto \mathbb{E}_{\mathbb{P}}\big[\psi_x(g, \boldsymbol{p})\big]$ is a contraction mapping. With $(\partial V(x), |\cdot|)$ a complete metric space, by Banach's fixed-point theorem, there is a unique $g(x) \in \partial V(x)$ satisfying $g(x) = \mathbb{E}_{\mathbb{P}}\big[\psi_x(g(x), \boldsymbol{p})\big]$. So (25) yields

$$\begin{aligned}
g(x) &= \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p^\star(x)\}\big]\big(\beta g(x) - c(x)\big) + \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}\big] \\
&= \frac{\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}\big] - \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p^\star(x)\}\big]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p^\star(x)\}\big]} \in \partial V(x).
\end{aligned} \tag{26}$$

Using this supergradient and observing that p is by definition the minimum acceptable price $p^\star(y(\infty, p))$, a first-order optimality condition for the unconstrained version of (21) is

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

which can be rearranged to

$$\begin{aligned}
c(y) &= -\big(1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p\}\big]\big)p + \beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p\}\big] \\
&= \beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p\}\big] - \beta\mathbb{E}_{\mathbb{P}}\big[p\,\mathbb{1}\{\boldsymbol{p} > p\}\big] - \big(1 - \beta\big)p \\
&= \beta\mathbb{E}_{\mathbb{P}}\big[(\boldsymbol{p} - p)_+\big] - (1 - \beta)p.
\end{aligned} \tag{27}$$

If this is an element of $[c(0), c(x)]$, then $c^{-1}\big(\beta\mathbb{E}_{\mathbb{P}}\big[(\boldsymbol{p} - p)_+\big] - (1 - \beta)p\big)$ solves (21). Otherwise, due to the concavity of the univariate objective function, first projecting (27) onto $[c(0), c(x)]$ solves (21). $\quad\square$

Appendix F: Proof of Proposition 3

We prove the result using several lemmas which provide formulae for $\frac{d}{dx}\widetilde{V}_\mathrm{S}(x)$ and $\frac{d}{dx}\widetilde{V}_\mathrm{M}(x)$. To this end the following lemma will be useful.

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 $F(x) := \mathbb{E}\big[f(x, \boldsymbol{\omega})\big]$. 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 $l(\boldsymbol{\omega})$ such that $\mathbb{E}\big[l(\boldsymbol{\omega})\big] < \infty$, and for each x, x' in a neighbourhood of $\bar{x}$ and almost every $\omega \in \Omega$ the following inequality holds:*

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

(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

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

Now the derivative values can be established.

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

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

$$\widetilde{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots) := \sum_{t=1}^\infty \beta^{t-1}\big(p^t(x^t - x^{t+1}) - C(x^{t+1})\big) \quad \text{where } x^{t+1} = y_{\mathrm{S}}(x^t, p^t) \, \forall t \in \mathbb{N}. \tag{28}$$

If $\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots$ are random price variables that are independent and identically distributed according to $\mathbb{P}$, the definition (8) shows that $\mathbb{E}_{\mathbb{P}^\infty}\big[\widetilde{v}_{\mathrm{S}}(x^1, \boldsymbol{p}^1, \boldsymbol{p}^2, \ldots)\big] = \widetilde{V}_{\mathrm{S}}(x^1)$. Recall that $p_{\mathrm{S}}^\star(x^1)$ 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 $p \geq p_{\mathrm{S}}^\star(x^1)$ the SDP policy sells down to the target inventory level $y_{\mathrm{S}}(\infty, p) \leq x^1$. It follows that in (28) all inventory levels $x^t = x^1$ for $t \leq T$, with T being the first time at which $p^T \geq p_{\mathrm{S}}^\star(x^1)$. 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 SDP policy. Under Assumption 3 the probability distribution $\mathbb{P}$ is atomless which implies that $\mathbb{P}\big[\boldsymbol{p}^t \neq p_{\mathrm{S}}^\star(x^1)\big] = 1$. It follows that in (28) there are no 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 $t \leq T$ 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 and we write

$$\frac{d\widetilde{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots)}{dx^1} = \beta^{T-1} p^T - \sum_{t=1}^{T-1} \beta^{t-1} c(x^1), \tag{29}$$

meeting (iii).

It remains to verify (ii). When attempting to evaluate $\mathbb{E}_{\mathbb{P}^\infty}\big[\frac{d}{dx^1}\widetilde{v}_{\mathrm{S}}(x^1, \boldsymbol{p}^1, \boldsymbol{p}^2, \ldots)\big]$ using the expression (29), the time $\boldsymbol{T}$ and the price $\boldsymbol{p^T}$ are random variables. The expectation of the first term in (29) is

$$\mathbb{E}_{\mathbb{P}^\infty}\big[\beta^{\boldsymbol{T}-1}\boldsymbol{p^T}\big] \leq \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\big|\boldsymbol{p} \geq p_{\mathrm{S}}^\star(x^1)\big] \leq \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\big|\boldsymbol{p} \geq \max\{p_1, \ldots, p_N\}\big] < \infty,$$

where the first inequality follows since $\boldsymbol{p^T}$ is the first price greater than $p_{\mathrm{S}}^\star(x^1)$, 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 $\mathbb{P}$ has a finite mean. The expectation of the second term in (29) is $\mathbb{E}_{\mathbb{P}^\infty}\big[\sum_{t=1}^{\boldsymbol{T}-1} \beta^{t-1} c(x^1)\big] \leq \frac{1}{1-\beta} c(x^1)$ which is bounded on compact sets. Together these observations show that $\widetilde{v}_{\mathrm{S}}(x^1, p^1, p^2, \ldots)$ has a Lipschitz constant in its first argument with finite expectation, meeting (ii). Thus, Lemma 2 applies and $\widetilde{V}_{\mathrm{S}}(x)$ is differentiable. Similar reasoning shows $\widetilde{V}_{\mathrm{M}}(x)$ is differentiable. $\square$

LEMMA 4. *Let Assumption 3 hold. Then the derivatives of the functions $\widetilde{V}_{\mathrm{S}}(x)$ and $\widetilde{V}_{\mathrm{M}}(x)$ are given by*

$$\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx} = \frac{\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{S}}^\star(x)\}\big] - \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \leq p_{\mathrm{S}}^\star(x)\}\big]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \leq p_{\mathrm{S}}^\star(x)\}\big]}$$

and

$$\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} = \frac{\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{M}}^\star(x)\}\big] - \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \leq p_{\mathrm{M}}^\star(x)\}\big]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \leq p_{\mathrm{M}}^\star(x)\}\big]},$$

respectively.

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

$$\widetilde{v}_{\mathrm{S}}(x, p) := \begin{cases} -C(x) + \beta\widetilde{V}_{\mathrm{S}}(x) & \text{if } p \leq p_{\mathrm{S}}^\star(x) \\ p\big(x - y_{\mathrm{S}}(\infty, p)\big) - C\big(y_{\mathrm{S}}(\infty, p)\big) + \beta\widetilde{V}_{\mathrm{S}}\big(y_{\mathrm{S}}(\infty, p)\big) & \text{if } p > p_{\mathrm{S}}^\star(x). \end{cases} \tag{30}$$

The equation (14) shows $\mathbb{E}_{\mathbb{P}}\big[\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})\big] = \widetilde{V}_{\mathrm{S}}(x)$. Recalling that $p_{\mathrm{S}}^{\star}$ is continuous, for each $p < p_{\mathrm{S}}^{\star}(x)$ there is a neighbourhood of x values for which the first case of (30) holds. Due to Lemma 3, $\frac{d}{dx}\widetilde{V}_{\mathrm{S}}(x)$ exists, and thus

$$\frac{d\widetilde{v}_{\mathrm{S}}(x,p)}{dx} = -c(x) + \beta\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx}. \tag{31}$$

In the other case, for each $p > p_{\mathrm{S}}^{\star}(x)$ the second case of (30) holds and

$$\frac{d\widetilde{v}_{\mathrm{S}}(x,p)}{dx} = p. \tag{32}$$

Similar reasoning to that in the proof of Lemma 3 shows $\frac{d}{dx}\widetilde{V}_{\mathrm{S}}(x) = \mathbb{E}_{\mathbb{P}}\big[\frac{d}{dx}\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})\big]$, and with $\mathbb{P}$ atomless, the event $\boldsymbol{p} = p_{\mathrm{S}}^{\star}(x)$ is immaterial when evaluating $\mathbb{E}_{\mathbb{P}}\big[\frac{d}{dx}\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})\big]$. Taking the expectation of (31) and (32) yields

$$\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx} = \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\big\{\boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\big\}\Big]\left(\beta\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx} - c(x)\right) + \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{\boldsymbol{p} > p_{\mathrm{S}}^{\star}(x)\big\}\Big]$$

$$= \frac{\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{S}}^{\star}(x)\}\big] - \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\}\big]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\}\big]}.$$

The expression for $\frac{d}{dx}\widetilde{V}_{\mathrm{M}}(x)$ can be derived in the same way. $\square$

Now Proposition 3 can be established.

PROPOSITION 3. *Let Assumption 3 hold. For price samples $p_1,\ldots,p_N$ which determine the sample-based SDP and MPC policies by (12) and (13), if the minimum acceptable SDP price $p_{\mathrm{S}}^{\star}(x;p_1,\ldots,p_N)$ is such that*

$$c(x) \ge \beta\mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{\boldsymbol{p} > p_{\mathrm{S}}^{\star}(x;p_1,\ldots,p_N)\big\}\Big] \quad \forall x \in [0,x^1],$$

then $\widetilde{V}_{\mathrm{M}}(x;p_1,\ldots,p_N) \ge \widetilde{V}_{\mathrm{S}}(x;p_1,\ldots,p_N)$ for each $x \in [0,x^1]$. That is, MPC performs at least as well out of sample on (ROSP) as SDP.

Proof of Proposition 3. Starting from initial inventory $x \in [0,x^1]$, let $\widetilde{V}_{\mathrm{S,M}}(x)$ denote the objective value of (ROSP) when following the SDP policy in the first stage and the MPC policy thereafter (as for Definition 2). We first compare derivatives and show $\frac{d}{dx}\widetilde{V}_{\mathrm{M}}(x) \ge \frac{d}{dx}\widetilde{V}_{\mathrm{S,M}}(x)$. Rearranging the identity provided by Lemma 4,

$$\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} = \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\big\{\boldsymbol{p} \le p_{\mathrm{M}}^{\star}(x)\big\}\Big]\left(\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x)\right) + \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{\boldsymbol{p} > p_{\mathrm{M}}^{\star}(x)\big\}\Big], \tag{33}$$

and similarly

$$\frac{d\widetilde{V}_{\mathrm{S,M}}(x)}{dx} = \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\big\{\boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\big\}\Big]\left(\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x)\right) + \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{\boldsymbol{p} > p_{\mathrm{S}}^{\star}(x)\big\}\Big].$$

Thus, to compare the derivatives it suffices to show

$$\mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\big\{p_{\mathrm{M}}^{\star}(x) < \boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\big\}\Big] \ge \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\big\{p_{\mathrm{M}}^{\star}(x) < \boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\big\}\Big]\left(\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x)\right). \tag{34}$$

From (33) we have

$$\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x) = \frac{\beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p_{\mathrm{M}}^{\star}(x)\}\big] - c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p_{\mathrm{M}}^{\star}(x)\}\big]} \le \frac{\beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{p_{\mathrm{M}}^{\star}(x) < \boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\}\big]}{1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p_{\mathrm{M}}^{\star}(x)\}\big]},$$

where the inequality follows from the condition in the statement of the proposition. With $\beta \in (0,1)$ and $1 - \beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p_{\mathrm{M}}^{\star}(x)\}\big] \ge \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{p_{\mathrm{M}}^{\star}(x) < \boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\}\big]$, multiplying by $\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{p_{\mathrm{M}}^{\star}(x) < \boldsymbol{p} \le p_{\mathrm{S}}^{\star}(x)\}\big]$ yields (34), whereby $\frac{d}{dx}\widetilde{V}_{\mathrm{M}}(x) \ge \frac{d}{dx}\widetilde{V}_{\mathrm{S,M}}(x)$. Hence, $\widetilde{V}_{\mathrm{M}}(x) \ge \widetilde{V}_{\mathrm{S,M}}(x)$ for all $x \in [0,x^1]$, and from Theorem 3 we conclude $\widetilde{V}_{\mathrm{M}}(x) \ge \widetilde{V}_{\mathrm{S}}(x)$ for each $x \in [0,x^1]$, as required. $\square$

Appendix G: **Proof of Proposition 4**

To make the sample-dependence explicit, we write $\widetilde{V}_S(x; p_1, \ldots, p_N) = \widetilde{V}_S(x)$ and $\widetilde{V}_M(x; p_1, \ldots, p_N) = \widetilde{V}_M(x)$. The expectations $\mathbb{E}_{\mathbb{P}^N}\left[\widetilde{V}_S(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\right]$ and $\mathbb{E}_{\mathbb{P}^N}\left[\widetilde{V}_M(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\right]$ are then the expected out-of-sample performances of the sample-based policies.

PROPOSITION 4 *Let $\mathbb{P}$ be the* 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 SDP on (ROSP) is unbounded below, while that of MPC is bounded.*

Proof of Proposition 4   First consider the term $\mathbb{E}_{\mathbb{P}^N}\left[\widetilde{V}_S(1; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\right]$. With $\widetilde{V}_S(0; p_1, \ldots, p_N) = 0$ we write $\widetilde{V}_S(1; p_1, \ldots, p_N) = \int_0^1 \frac{d}{dx}\widetilde{V}_S(x; p_1, \ldots, p_N)\, dx$ using the expression for the derivative given by Lemma 4. Hence,

$$\mathbb{E}_{\mathbb{P}^N}\left[\widetilde{V}_S(1; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\right] = \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)}^\infty p f(p)\, dp - x F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}{1 - \beta F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}\, dx\right], \quad (35)$$

where F and f denote the cumulative distribution function and probability density function of the Exponential(1) distribution, respectively. The negative term in (35) is

$$\mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{-x F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}{1 - \beta F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}\, dx\right] =$$

$$\int_0^\infty \cdots \int_0^\infty \int_0^1 \frac{-x F(p_S^\star(x; p_1, \ldots, p_N))}{1 - \beta F(p_S^\star(x; p_1, \ldots, p_N))}\, dx\, f(p_N)\, dp_N \cdots f(p_1)\, dp_1. \quad (36)$$

Fix $p_1, \ldots, p_{N-1}$ and consider the inner-most integral in (36) when p_N is large. Following (12), for $\alpha \in (0, 1)$ and all $\beta \in [\alpha, 1)$, $x \in [0, 1]$, the value $p_S^\star(x; p_1, \ldots, p_N)$ is the p which solves $\beta \frac{1}{N}(p_N - p)_+ + (1 - \beta)p = x$. Thus

$$\frac{-x F(p_S^\star(x; p_1, \ldots, p_N))}{1 - \beta F(p_S^\star(x; p_1, \ldots, p_N))} \to \frac{-x(1 - e^{-(p_N - Nx)})}{1 - (1 - e^{-(p_N - Nx)})}$$

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

$$\lim_{\beta \to 1} \int_0^1 \frac{-x F(p_S^\star(x; p_1, \ldots, p_N))}{1 - \beta F(p_S^\star(x; p_1, \ldots, p_N))}\, dx = \int_0^1 \frac{-x(1 - e^{-(p_N - Nx)})}{1 - (1 - e^{-(p_N - Nx)})}\, dx = \frac{1}{2} + \left(\frac{1}{N} + \frac{1}{N^2}\right)e^{p_N - N} - \frac{1}{N^2}e^{p_N}.$$

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

$$\int_0^1 \frac{-x F(p_S^\star(x; p_1, \ldots, p_N))}{1 - \beta F(p_S^\star(x; p_1, \ldots, p_N))}\, dx < \frac{1}{2} + \left(\frac{1}{N} + \frac{1}{N^2}\right)e^{p_N - N} - \frac{1}{N^2}e^{p_N} + \varepsilon. \quad (37)$$

But

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

and for $0 < a < b$ the integral

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

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

$$\int_0^\infty \int_0^1 \frac{-x F(p_S^\star(x; p_1, \ldots, p_N))}{1 - \beta F(p_S^\star(x; p_1, \ldots, p_N))}\, dx\, f(p_N)\, dp_N$$

can be made arbitrarily negative by increasing β towards 1, regardless of the values of $p_1, \ldots, p_{N-1}$. We therefore conclude that (36) is unbounded below as $\beta \to 1$.

Recall that $p_S^\star(x; p_1, \ldots, p_N) \leq \max\{p_1, \ldots, p_N\}$. The positive term in (35) is

$$\mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)}^\infty p f(p)\, dp}{1 - \beta F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}\, dx\right] \leq \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)}^\infty p f(p)\, dp}{1 - F(p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N))}\, dx\right] \qquad (38)$$

$$= \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 1 + p_S^\star(x; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\, dx\right]$$

$$\leq \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 1 + \max\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N\}\, dx\right]$$

$$= \mathbb{E}_{\mathbb{P}^N}\left[1 + \max\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N\}\right].$$

Here $\mathbb{E}_{\mathbb{P}^N}\left[\max\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N\}\right]$ is the expected value of the Nth order statistic of N random samples from the Exponential(1) distribution, which is finite; see, e.g., (David and Nagaraja 2004, Section 4.2). Hence, (38) is bounded above for all $\beta \in (0, 1)$. We have already shown that (36) is unbounded below as $\beta \to 1$. Thus, the sum (35) is unbounded below as $\beta \to 1$, regardless of the choice of N.

Consider now $\mathbb{E}_{\mathbb{P}^N}\left[\widetilde{V}_M(1; \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)\right]$. For simplicity, we set $N = 2$, although the result can be shown to hold for $N \geq 2$. As for (35),

$$\mathbb{E}_{\mathbb{P}^2}\left[\widetilde{V}_M(1; \boldsymbol{p}_1, \boldsymbol{p}_2)\right] = \mathbb{E}_{\mathbb{P}^2}\left[\int_0^1 \frac{\int_{p_M^\star(x; \boldsymbol{p}_1, \boldsymbol{p}_2)}^\infty p f(p)\, dp - x F(p_M^\star(x; \boldsymbol{p}_1, \boldsymbol{p}_2))}{1 - \beta F(p_M^\star(x; \boldsymbol{p}_1, \boldsymbol{p}_2))}\, dx\right], \qquad (39)$$

which has negative term

$$\mathbb{E}_{\mathbb{P}^2}\left[\int_0^1 \frac{-x F(p_M^\star(x; \boldsymbol{p}_1, \boldsymbol{p}_2))}{1 - \beta F(p_M^\star(x; \boldsymbol{p}_1, \boldsymbol{p}_2))}\, dx\right] = \int_0^\infty \int_0^\infty \int_0^1 \frac{-x F(p_M^\star(x; p_1, p_2))}{1 - \beta F(p_M^\star(x; p_1, p_2))}\, dx\, f(p_2)\, dp_2\, f(p_1)\, dp_1. \qquad (40)$$

The iterated integral (40) can be divided into ranges based on the value of the sample average $\mu_2 = \frac{1}{2}(p_1 + p_2)$. Following (13), depending on the value of μ_2, either $p_M^\star(x; p_1, p_2) = \beta\mu_2 - x$ or $p_M^\star(x; p_1, p_2) = -x/(1 - \beta)$. Since $\mu_2 - x \geq \beta\mu_2 - x \geq -x/(1 - \beta)$,

$$\int_0^1 \frac{-x F(p_M^\star(x; p_1, p_2))}{1 - \beta F(p_M^\star(x; p_1, p_2))}\, dx \geq \int_0^1 \frac{-x F(\mu_2 - x)}{1 - F(\mu_2 - x)}\, dx$$

$$= \int_0^{\min\{\mu_2, 1\}} \frac{-x(1 - e^{-(\mu_2 - x)})}{1 - (1 - e^{-(\mu_2 - x)})}\, dx$$

$$= \frac{1}{2}(\min\{\mu_2, 1\})^2 + (1 + \min\{\mu_2, 1\})e^{\mu_2 - \min\{\mu_2, 1\}} - e^{\mu_2}$$

$$\geq -e^{\mu_2}.$$

Thus, (40) is bounded below by

$$\int_0^\infty \int_0^\infty -e^{\frac{1}{2}(p_1 + p_2)} e^{-p_2}\, dp_2\, e^{-p_1}\, dp_1 = -4$$

for all $\beta \in (0, 1)$. Similar reasoning as for (38) shows that the positive term in (39) is bounded above. We therefore conclude that the sum (39) is bounded as $\beta \to 1$. $\quad\square$

Acknowledgments

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, 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), third 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.

Dowson O (2018) *Applying Stochastic Optimisation to the New Zealand Dairy Industry*. PhD thesis, University of Auckland.

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 MVF, Pinto LMVG (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: SIAM).

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 Jr RE, 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.