Stochastic Control and Optimization with Nonstationary Data

Dominic Samuel Tecumseh Keehan

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Operations Research,
the University of Auckland, 2025.

Abstract

Data-driven decision making seeks to learn from the past how to operate in an uncertain future. In this thesis we aim to understand what drives the performance of, and thus how to make, effective data-driven decisions.

In the first part of the thesis, we assume that the data are stationary and provide samples from the true distribution governing uncertainty. We focus on stochastic optimal control problems and decisions derived from approximations of associated stochastic dynamic programming functional equations. We consider approximation accuracy, finding that epi-consistency depends upon the existence of sets of functions under which the approximating Bellman operators satisfy asymptotic integrability and semicontinuity conditions. Generality here allows us to treat the simultaneous approximation of objective functions and constraints. We then study out-of-sample performance through a distributionally robust optimization lens, showing that sample-based model predictive control can outperform stochastic dynamic programming when the Bellman operators preserve convexity — a characteristic of problems with uncertainty appearing in the objective function. Conversely, with uncertainty instead appearing in the constraints, stochastic dynamic programming can outperform model predictive control. We apply this understanding to a supply chain problem from the New Zealand dairy industry, improving decision and predictive performance by 11 % and 40 %, respectively.

In the second part of the thesis, we relax the assumption of stationarity and work with data from distributions that change over time. With distributions at the heart of data-driven approximation schemes, we focus on the estimation of nonstationary distributions via weighted empirical distributions. We derive Wasserstein-distance concentration bounds that reveal a fundamental variance–drift trade-off between the numerousness of the weighted data and our exposure to distributional drift. Optimally balancing this trade-off — from a minimally conservative distributionally robust optimization perspective — yields weightings which decay polynomially in time. We then propose a maximum likelihood–estimation approach, regularizing by the Wasserstein distance to control temporal smoothness. The resulting finite-dimensional convex exponential-cone program is applied to forecasting international dairy commodity prices, improving performance by 8 %.

Taken together, the two parts of this thesis lay the foundations for data-driven stochastic control in data-scarce and nonstationary environments.

To Alison and Michael

Acknowledgements

I would like to thank my supervisor Andy Philpott and my co-supervisor Andrew Mason, as well as Eddie Anderson, Oscar Dowson, Johannes Royset, Ryan Cory-Wright, and Wolfram Wiesemann. Your kindness, generosity of time, relentless positivity, view for generality, and intuition, have made this thesis possible. A special thanks to Arkadii Slinko for setting this all in motion. I would also like to thank the graduate students and the staff at Imperial Business School for their hospitality. Finally, I would like to thank my examiners, Daniel Kuhn and Rachael Tappenden, for their timely provision of examination reports and interesting questions.

I was financially supported by the University of Auckland Doctoral Scholarship, the Claude McCarthy Fellowship, and the Department of Engineering Science & Biomedical Engineering. I am very grateful.

Dom,

38,000 ft above the Pacific Ocean

Contents

Chapter 1

Introduction

Data-driven decision making seeks to learn from the past how to operate in an uncertain future. By analysing historical data, we aim to uncover patterns governing stochastic phenomena and construct models that guide our actions. The paradigm is central to modern operations, with applications in engineering, business, healthcare, and beyond. Examples abound: electric-power schedulers use historical inflows to coordinate reservoir levels and thermal-energy generation, supply chain managers use historical demand to determine restocking levels, and service-system planners use historical arrivals to design triaging strategies.

In this thesis we are interested in data-driven decision making. A classical example is the newsvendor problem, where a merchant selling a perishable good must decide how much product to purchase from a third-party supplier at the beginning of the day, in advance of customers arriving sporadically throughout the rest of the day. Purchasing too little product incurs a heavy opportunity cost from lost sales and diminished consumer confidence. Purchasing too much product destroys profit margins from the spoilage of unsold product. Assuming that the behaviour of customers is unchanging over time, once ample historical data (i.e., daily demand counts intimated from customer inquiries) has been collected, the merchant can begin to make informed approximations of the underlying demand *distribution*, and hence make purchasing decisions that optimally balance missed sales against unsold product.

Data often come in the form of discrete pieces of information, or *samples*. The most basic approach to decision making works with an empirical probability distribution on samples as a distributional approximation [76]. Although well-motivated asymptotically, due to a dearth of historical observations, or out of computational necessity, using a large number of samples is often not possible. This makes the empirical approach vulnerable to *overfitting*, where the model

conforms too closely to the data and fails to generalise. In the context of the aforementioned newsvendor problem, our merchant could be forced to make purchasing decisions using only a couple of weeks of historical demand counts, either due to having just recently opened up shop, expanding into a new product market, or recent changes to consumer purchasing behaviour. With a limited history of observations, the empirical distribution is unlikely to adequately represent the underlying demand distribution.

Sometimes referred to as the optimizer's curse, overfitting tends to yield out-of-sample disappointment: once the resulting model is optimized for a decision, the sample set leads to an optimistically biased estimator of the true out-of-sample performance [112]. Yet, for adoption in industry, it is essential that decisions can be trusted out of sample — if a decision-making policy performs significantly worse than predicted, or very poorly outright, the consequences can be severe. To that end, in this thesis we study different distributional approximations, with the goal of understanding what drives out-of-sample performance, mitigating the effects of overfitting, and extracting as much relevant information as possible from historical data.

The newsvendor problem we previously discussed is an example of a *single*-period planning problem. In many problems of great practical interest, our decisions and their consequences unfold sequentially over *multiple* time periods. In this thesis we will frequently consider a specific class of multistage (i.e., multi-period) data-driven decision making problems: that of a supply chain manager coordinating the allocation of periodic resource inflows into commodities ready for sale on a spot market. With spot prices evolving stochastically over time, the problem is what commodities to produce and when to sell them on the market. We will focus on different aspects of an application arising in the New Zealand dairy industry. Here, each month dairy companies use data to model the likely long-term behaviour of international market prices, thereby deciding what nonperishable dairy products to manufacture and when to export them. A good data-driven approximation must yield profits for the company and accurate advance forecasts of such profits for communication to owner-operator shareholders. See Figure 1.1 for the historical price data that New Zealand dairy companies grapple with when making these production and export decisions. (The figure presents the prices of five different dairy commodities; we defer a specific discussion of these commodities until later in the thesis.)

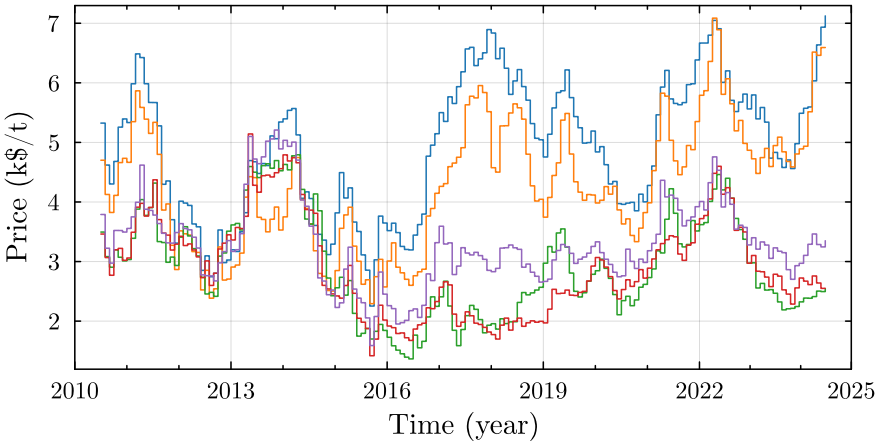

Figure 1.1: **Historical International Dairy Commodity Prices**[1]

Multiperiod decision making complicates the process of learning from data for two reasons. The first reason is computational in nature. Stochastic optimal control formalises the problem of decision-making over time, and here dynamic programming is a fundamental solution technique [18]. While powerful, dynamic programming suffers from the well-known curse of dimensionality, whereby problems scale poorly and can be computationally expensive to solve [13]. In addition to approximating uncertain distributions, approximating structural aspects of problems is usually necessary to facilitate efficient solution. Each born from different fields, the methods of stochastic dual dynamic programming [90], reinforcement learning [16], and model predictive control [17] (among many others), all approach this approximation differently. Whether or not one method is suited to a particular application depends on the interplay of problem characteristics and the relative importance of data-driven versus structural approximation.

The second reason concerns the relevance of data. In many applications, the underlying stochastic phenomena are unlikely to be stationary and will instead change over time. Within a *non*-stationary environment, due to the gradual accumulation of changes, historical observations provide less relevant information as they become further removed from the present. In this way, data-driven approximations must balance the possible statistical gains of using lengthy historical records against the need to respond to changes in the underlying stochasticity.

This thesis comes in two parts. In the first part we work to understand what drives the out-of-sample performance of approximating decision policies when the underlying sample-generating distribution is stationary. We focus on the approximation of stochastic optimal control problems through their associated stochastic dynamic programming functional equations. We consider (i) approximations arising from the data-driven aspects of modelling uncertainty,

[1] Data retrieved from [39].

and (ii) approximations that further simplify structural aspects of problems. Attention is paid to the asymptotic convergence of approximations, and to the effects of problem characteristics on out-of-sample performance. As practical illustrations of our results, we often apply them to components of the New Zealand dairy supply chain problem discussed earlier.

In the second part of this thesis we relax the assumption of stationarity and work to understand how to best approximate the distributions used to build decision policies. We focus on the approximation of nonstationary probability distributions through *weighted* empirical distributions. Attention is paid to weightings which guarantee that the resulting empirical distribution is close to the underlying distribution with high probability, and to how particular data outcomes reveal likely changes in the underlying distribution. We again apply this to the New Zealand dairy supply chain, addressing a price-forecasting component of the problem.

1.1 Background

We now provide some mathematical background to establish notation and provide a framework for the studies in this thesis. So that the content within each chapter is self-contained, some of this information is also recalled as necessary in the body of the thesis.

Notation

$\mathbb{R}$ denotes the set of real numbers and $\mathbb{N}$ denotes the set of natural numbers. For $N \in \mathbb{N}$, we write $[N] \coloneqq \{1, \ldots, N\}$. For $p \in [1, \infty]$, we write $\|\cdot\|_p$ for the L_p norm. For a metric space $(\mathcal{Z}, d_p)$, $p \in [1, \infty]$, the p-ball of radius ε centred at $z \in \mathcal{Z}$ is denoted $\mathbb{B}_p(z; \varepsilon) \coloneqq \{z' \in \mathcal{Z} : d_p(z, z') \leq \varepsilon\}$.

$\mathfrak{P}(\Xi)$ denotes the set of Borel probability measures on a closed set $\Xi \subseteq \mathbb{R}^m$. All of the distributions considered in this thesis are Borel probability measures. Among these distributions, the point-mass distribution $\mathbb{1}_\xi$ assigns probability 1 to the outcome $\xi \in \Xi$. For a distribution $\mathbb{P} \in \mathfrak{P}(\Xi)$, we write $\boldsymbol{\xi} \sim \mathbb{P}$ to express that a random vector $\boldsymbol{\xi}$ is distributed according to $\mathbb{P}$. (We use boldface to distinguish random vectors from their outcomes.) We write $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})] \coloneqq \int_\Xi f(\xi)\, d\mathbb{P}(\xi)$ for the expectation of a Borel-measurable function $f : \Xi \to \mathbb{R}$ when this is well defined.

For $x \in \mathbb{R} \cup \{-\infty, \infty\}$ and an interval $[a, b] \subseteq \mathbb{R} \cup \{-\infty, \infty\}$, we write $x_{[a,b]}$ to denote the projection of x onto $[a, b]$. For shorthand, we set $x_+ \coloneqq x_{[0,\infty]}$. We use $\mathbb{1}$ to denote the event indicator; that is, $\mathbb{1}\{\mathcal{E}\} = 1$ if the event $\mathcal{E}$ occurs and $\mathbb{1}\{\mathcal{E}\} = 0$ if it does not. With a slight

abuse of notation, occasionally we will apply an operation defined for a scalar to a vector or to a set; in such cases the operation is understood to be applied componentwise, or to all elements in the set, respectively.

Stochastic Optimal Control

Given a closed state space $\mathcal{X} \subseteq \mathbb{R}^n$ and outcome space $\Xi \subseteq \mathbb{R}^m$, we consider discrete-stage stochastic optimal control problems of the form

$$\begin{aligned} \underset{y_1,\dots,y_T}{\text{minimize}} \quad & \mathbb{E}_{\mathbb{P}^T}\left[\sum_{t=1}^{T} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \\ \text{subject to} \quad & x_{t+1} = y_t(x_t, \boldsymbol{\xi}_t) \in \mathcal{Y}(x_t, \boldsymbol{\xi}_t), \quad t \in [T], \end{aligned} \tag{SOC}$$

starting from an initial state $x_1 \in \mathcal{X}$. The horizon $T \in \mathbb{N} \cup \{\infty\}$ may be finite or infinite, and the stagewise-independent random vector $\boldsymbol{\xi}_t \sim \mathbb{P} \in \mathfrak{P}(\Xi)$ for each $t \in [T]$. Here the discount factor $\beta \in (0, 1)$, and the stage-cost function $\varphi : \mathcal{X} \times \mathcal{X} \times \Xi \to \mathbb{R}$. The set-valued mapping $\mathcal{Y} : \mathcal{X} \times \Xi \rightrightarrows \mathcal{X}$, assumed throughout to be nonempty and compact valued, defines admissible policies $\{y_1, \dots, y_T : \mathcal{X} \times \Xi \to \mathcal{X}\}$. For simplicity we do not vary $\mathbb{P}$, φ, or $\mathcal{Y}$ between stages, however, all of our eventual results can be appropriately generalised to such a setting. Naturally, the problem requires additional assumptions to be well defined. We will focus on the associated dynamic programming equations and return to the discussion of assumptions in the following chapters.

Example 1.1. As an example of (SOC), consider a supply chain manager coordinating the production and sales of m commodities over a planning horizon of T periods. At each time $t \in [T]$ the manager observes the current inventory of products $x_t \in \mathcal{X} \subseteq \mathbb{R}^m_+$ and the current spot price vector $p_t \in \mathbb{R}^m_+$. The future spot price vector $\boldsymbol{p}_{t+1}$ is *random*. The manager then decides how much to sell $u_t \in \mathcal{U}(x_t, p_t) \subseteq \mathbb{R}^m_+$ and how much to produce $w_t \in \mathcal{W}(x_t, p_t) \subseteq \mathbb{R}^m_+$, subject to inventory-storage and production cost functions C and $H : \mathcal{X} \to \mathbb{R}$. The objective is to maximize

the expected total discounted future profit:

$$\begin{aligned}
\underset{\substack{u_1,\dots,u_T\\ w_1,\dots,w_T}}{\text{maximize}} \quad & \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} \cdot \Big(\boldsymbol{p}_t^\intercal u_t(x_t, \boldsymbol{p}_t) - C\big(x_t - u_t(x_t, \boldsymbol{p}_t)\big) - H\big(w_t(x_t, \boldsymbol{p}_t)\big)\Big)\right], \\
\text{subject to} \quad & x_{t+1} = x_t - u_t(x_t, \boldsymbol{p}_t) + w_t(x_t, \boldsymbol{p}_t) \in \mathcal{X}, \qquad t \in [T], \\
& u_t(x_t, \boldsymbol{p}_t) \in \mathcal{U}(x_t, \boldsymbol{p}_t), \qquad t \in [T], \\
& w_t(x_t, \boldsymbol{p}_t) \in \mathcal{W}(x_t, \boldsymbol{p}_t), \qquad t \in [T].
\end{aligned}$$

This constitutes a problem where the supply chain manager faces two trade-offs: first, allocating limited resource inflows among the production of different commodities; and second, holding inventory to exploit the possibility of higher spot prices in the future versus selling to secure revenue immediately.

So far we have been intentionally vague about the distribution of the spot prices. There are many possibilities. Modelling the prices as timewise independent and identically distributed, one could set $\boldsymbol{p}_t \sim \mathbb{P} \in \mathfrak{P}(\mathbb{R}^m_+)$. More generally, modelling the prices as evolving over time according to a lag-one vector-autoregressive process with parameters $\mu \in \mathbb{R}^m_+$ and $A \in \mathbb{R}^{m\times m}$, one could set $\boldsymbol{p}_t = \mu + Ap_{t-1} + \boldsymbol{\xi}_t$, where each $\boldsymbol{\xi}_t \sim \mathbb{P} \in \mathfrak{P}(\mathbb{R}^m)$ is a random innovation term. To capture multiplicative effects in the evolution of spot prices, we will often use an autoregressive model that sets $\log(\boldsymbol{p}_t) = \mu + A\log(p_{t-1}) + \boldsymbol{\xi}_t$. □

A problem closely related to (SOC) is that of finding a family of real-valued solution functions $v_t : \mathcal{X} \times \Xi \to \mathbb{R}$, $t \in [T]$, to the stochastic dynamic programming (SDP) functional equations

$$v_t(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{\varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}}\big[v_{t+1}(y, \boldsymbol{\xi})\big]\Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi,\ t \in [T], \tag{1.1}$$

where $v_{T+1} \coloneqq 0$. Given such v_t, $t \in [T]$, a function $y_t : \mathcal{X} \times \Xi \to \mathcal{X}$ with

$$y_t(x, \xi) \in \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \Big\{\varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}}\big[v_{t+1}(y, \boldsymbol{\xi})\big]\Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi,$$

and with $\xi \mapsto y_t(x, \xi)$ measurable for all $x \in \mathcal{X}$, defines a decision rule at stage t. Repeating this for each stage defines an admissible policy for the control problem. If $\mathbb{E}_{\mathbb{P}}\big[v_t(\,\cdot\,, \boldsymbol{\xi})\big]$, $t \in [T]$, satisfies the principle of optimality, then such a policy solves (SOC) and the value $\mathbb{E}_{\mathbb{P}}\big[v_1(x_1, \boldsymbol{\xi})\big]$ is equal to its optimal objective value; see, e.g., [113, Theorem 9.2].

Reformulating (SOC) as finding a solution to (1.1) opens up key avenues for analysis. Accordingly, the principle of optimality is fundamental to this thesis, and we will mobilise appropriate results to ensure that it holds in a range of different contexts. The simplest results require φ to be bounded and continuous, or to satisfy a certain monotonicity property. For a detailed discussion of these conditions, see, e.g., [18, Propositions 3.1 and 3.2] in the finite-horizon case; and [113, Theorem 9.6] and [18, Propositions 4.2, 5.2, and 5.3] in the infinite-horizon case.

Inherent in the problem (SOC) is the distribution $\mathbb{P}$ which is our *distributional* forecast for the behaviour of the random variables in future stages. As a result, (SOC) and (1.1) are computationally expensive to solve in general. Instead of working directly with $\mathbb{P}$, model predictive control (MPC) simplifies the problem by working with a *deterministic* forecast. This forecast could be generated as the sequence of iterates of a linear recurrence with tuned coefficients, or from a more specialised machine learning algorithm. In this thesis we will view MPC as working with a deterministic forecast $\mu \in \Xi$ of the future random variables in every stage. (Note that this still permits the case of a linear recurrence by reformulating the random variables as the consecutive changes in a noisy state variable, e.g., the innovations in an autoregression.)

MPC is often understood to be applied in a "rolling-horizon" fashion, where the problem is solved at the current state and outcome pair (x, ξ), the resulting optimal decision is enacted, we roll forward to the next stage with a new state and outcome pair (x', ξ'), and a new forecast is generated to then repeat the process. This is in contrast to dynamic programming, which computes a policy that provides a decision rule to follow at all possible state and outcome pairs in all future stages. However, if the randomness is stagewise independent, then MPC fits into the same functional-analytic framework as SDP, and we end up with the problem of finding a family of real-valued solution functions $v_t : \mathcal{X} \times \Xi \to \mathbb{R}$, $t \in [T]$, to the MPC functional equations

$$v_t(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta v_{t+1}(y, \mu) \Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi,\ t \in [T], \tag{1.2}$$

where $v_{T+1} \coloneqq 0$. Just as for SDP, if measurable functions which provide decisions that attain the infima in (1.2) can be found, then one has access to an admissible policy for (SOC).

Although somewhat naïve from a conceptual perspective, for a careful choice of forecast that distils the relevance of $\mathbb{P}$ to (SOC) down to its essential elements, MPC policies can sometimes perform quite well. A common (though not necessarily careful!) choice is to set $\mu \coloneqq \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$. Most importantly, finding solutions to (1.2) is usually much easier computationally than it is for (1.1).

Data-Driven Optimization

For an uncertainty-afflicted cost function $f : \mathcal{Z} \times \Xi \to \mathbb{R}$ and an unknown probability distribution $\mathbb{P} \in \mathfrak{P}(\Xi)$, consider the decision problem

$$\underset{z \in \mathcal{Z}}{\text{minimize}} \quad \mathbb{E}_{\mathbb{P}}[f(z, \boldsymbol{\xi})]. \tag{1.3}$$

This is the general form of both (SOC) and the minimizations in the functional equations (1.1). In lieu of the true distribution $\mathbb{P}$, using data we can estimate a surrogate distribution $\mathbb{Q} \in \mathfrak{P}(\Xi)$ and instead solve

$$\underset{z \in \mathcal{Z}}{\text{minimize}} \quad \mathbb{E}_{\mathbb{Q}}[f(z, \boldsymbol{\xi})] \tag{1.4}$$

to approximate (1.3). The most basic approach, termed *sample average approximation* (*or* SAA), uses N samples $\xi_1, \dots, \xi_N \in \Xi$ to estimate the surrogate empirical distribution $\mathbb{P}_N \coloneqq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\xi_i}$. For instance, in the context of our supply chain problem from Example 1.1, if spot prices are modelled as timewise independent, then given historical prices $p_1, \dots, p_N \in \mathbb{R}^m_+$, the SAA approach would approximate the distributions of the future price vectors as $\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{p_i}$. Or, if spot prices are modelled as evolving over time according to some lag-one vector-autoregressive process, then using estimates $\mu \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times m}$ of the underlying autoregression parameters, one could compute the implied historical innovations $\xi_i \coloneqq p_{i+1} - (\mu + Ap_i) \in \mathbb{R}^m$ for each consecutive historical-price pair (p_i, p_{i+1}), and the SAA approach would approximate the innovation distribution as $\frac{1}{N-1}\sum_{i=1}^{N-1} \mathbb{1}_{\xi_i}$.

One motivation for SAA is the strong law of large numbers: for broad classes of functions $g : \Xi \to \mathbb{R}$ and probability distributions $\mathbb{P} \in \mathfrak{P}(\Xi)$, given independent random samples $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_N \sim \mathbb{P}$ the empirical expectation $\mathbb{E}_{\mathbb{P}_N}[g(\boldsymbol{\xi})] = \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{\xi}_i) \to \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{\xi})]$ almost surely as $N \to \infty$; see, e.g., [21, Theorem 22.1]. Suitable *uniform* strong laws of large numbers further imply the following consistency result:

$$\text{if } z_N \in \operatorname*{argmin}_{z \in \mathcal{Z}} \Big\{ \mathbb{E}_{\mathbb{P}_N}[f(z, \boldsymbol{\xi})] \Big\} \to z^\star \in \mathcal{Z}, \quad \text{then } z^\star \in \operatorname*{argmin}_{z \in \mathcal{Z}} \Big\{ \mathbb{E}_{\mathbb{P}}[f(z, \boldsymbol{\xi})] \Big\} \text{ almost surely;}$$

see [74] and [96, Theorem 5.5(b)]. In other words, convergent sequences of approximating decisions obtained from (1.4) will converge to an optimal decision for the true problem (1.3).

Note that the aforementioned optimality is only valid asymptotically, and for a finite number of samples there are many attractive alternatives to SAA. One might use parametric information to

improve the estimation of the underlying distribution through a maximum-likelihood approach [4]. Moreover, for a finite number of samples, hinging decisions solely on an empirical distribution risks overfitting to that sample set. In general the true distribution $\mathbb{P}$ is an infinite-dimensional object, and perturbations to any one of the samples $\xi_1, \ldots, \xi_N$ will alter the approximating decision yielded from (1.4). Who is to say whether or not the particular outcome of data is representative of the complexities of $\mathbb{P}$? Instead of working with a single distributional estimate, *distributionally robust optimization* (*or* DRO) attempts to protect against perturbations to the data by optimizing conservatively over an *ambiguity set* of distributions $\mathcal{P}_N \subseteq \mathfrak{P}(\Xi)$ informed by the samples $\xi_1, \ldots, \xi_N$, thereby solving

$$\underset{z \in \mathcal{Z}}{\text{minimize}} \quad \sup_{\mathbb{Q} \in \mathcal{P}_N} \mathbb{E}_{\mathbb{Q}}[f(z, \boldsymbol{\xi})]. \tag{1.5}$$

In this way we now hope to estimate a *set* of distributions containing $\mathbb{P}$ with high probability; the construction of $\mathcal{P}_N$ and how we can reasonably ensure that the event $\mathbb{P} \in \mathcal{P}_N$ occurs is the basis for powerful out-of-sample performance guarantees [79].

For certain choices of ambiguity set $\mathcal{P}_N$, the min–sup problem (1.5) can often be reformulated as a tractable convex optimization problem. As such, in this thesis we will work with moment-based ambiguity sets and their generalisations. For example, a mean-based ambiguity set takes the form

$$\mathcal{P}_N = \Big\{ \mathbb{Q} \in \mathfrak{P}(\Xi) : \mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mathbb{E}_{\mathbb{P}_N}[\boldsymbol{\xi}] \Big\},$$

consisting of all of the distributions with mean vectors that match the empirical $\mathbb{E}_{\mathbb{P}_N}[\boldsymbol{\xi}]$.

Although adequate for some problems, moment-based ambiguity sets may not be sufficiently expressive, and hence can be overly conservative. We will thus also consider ambiguity sets based on the Wasserstein distances. For a Polish metric space (Ξ, d) and $p \in [1, \infty]$, the *p-Wasserstein distance* between two distributions $\mathbb{P}, \mathbb{Q} \in \mathfrak{P}(\Xi)$ is

$$W_p(\mathbb{P}, \mathbb{Q}) \coloneqq \Big(\inf_{\gamma \in \Gamma(\mathbb{P}, \mathbb{Q})} \mathbb{E}_{\gamma}\big[d(\boldsymbol{\xi}, \boldsymbol{\zeta})^p \big] \Big)^{1/p},$$

where $\Gamma(\mathbb{P}, \mathbb{Q})$ is the set of all couplings of $\mathbb{P}$ and $\mathbb{Q}$; that is, the set of all probability distributions on $\Xi \times \Xi$ with first marginal $\mathbb{P}$ and second marginal $\mathbb{Q}$. We assume throughout the thesis that the distance d is lower semicontinuous which implies that the infimum is attained [118, Theorem 4.1]. The Wasserstein distances metrise the weak convergence of probability measures, and by varying

the metric d and the order p, they provide the decision maker with a flexible distributional modelling tool. We will use p-Wasserstein balls as ambiguity sets, taking

$$\mathcal{P}_N = \mathbb{B}_p(\mathbb{P}_N; \varepsilon) \coloneqq \Big\{\mathbb{Q} \in \mathfrak{P}(\Xi) : W_p(\mathbb{P}_N, \mathbb{Q}) \leq \varepsilon\Big\}.$$

Here the radius of the ambiguity ball ε introduces a bias–variance trade-off: larger values yield decisions that are more biased but tend to exhibit lower variance and more reliable performance out of sample, while smaller values reduce bias but may lead to overfitting and poor generalisation out of sample. To balance these effects, ε is then a parameter that the decision maker can calibrate based on problem context and the available data.

Contributions

The contributions of the chapters in Part I of this thesis are as follows:

(i) **Epi-convergence of sample-based stochastic dynamic programs.** In Chapter 2 we provide general conditions under which sample-based approximations of stochastic dynamic programs asymptotically epi-converge to the true problem. We also give examples where these conditions hold and convergence is assured, as well as an example where these conditions do not hold and convergence fails.

(ii) **Comparing SDP and MPC through distributional robustness.** In Chapter 3 we provide conditions under which MPC can be interpreted as a mean-constrained distributionally robust version of sample-based SDP. These conditions provide a structural divider between the types of problems for which SDP and MPC are each likely to be suited. We then focus on a specific revenue optimization problem and provide conditions under which MPC outperforms SDP out of sample.

(iii) **Applying MPC to the New Zealand dairy supply chain.** In Chapter 4 we apply MPC to a supply chain–management problem arising in the New Zealand dairy industry. We highlight similarities to the revenue optimization problem of Chapter 3, and then show experimentally that MPC significantly outperforms other solution approaches when evaluated out of sample.

The contributions of the chapters in Part II of this thesis are as follows:

(i) **Weighted empirical distributions in nonstationary settings.** In Chapter 5 we provide Wasserstein-distance concentration bounds for weighted empirical distributions in nonstationary environments. We then optimize the associated performance guarantees in these weights. This reveals the worst-case decay rate governing the relevance of historical samples to Wasserstein distributionally robust optimization problems.

(ii) **Estimating distribution sequences via Wasserstein regularization.** In Chapter 6 we formulate an estimation problem for the most-likely sequence of distributions underlying a sequence of historical samples. By incorporating a Wasserstein-distance regularization term, this problem allows us to flexibly balance the relevance of the information contained in the outcomes of historical samples against its decay over time. The estimation problem becomes a solvable finite-dimensional convex-cone program. We then show experimentally that this estimation approach outperforms other approaches to a price-forecasting component of the dairy supply chain problem from Chapter 4.

Part I

Stochastic Optimal Control

Chapter 2

Consistent Approximation of Stochastic Dynamic Programs

In this chapter we consider stochastic optimal control problems where the solutions of the associated stochastic dynamic programming (SDP) equations are approximated. Towards our goal of understanding what drives the performance of different data-driven approximations, as a prerequisite to studying *out-of-sample* performance, here we focus on asymptotic convergence *in sample*. For instance, recall that sample average approximation (or SAA) uses the empirical distribution over a sample set as a distributional approximation. Assuming that these samples are drawn from the true underlying probability distribution, it is conceptually important that the resulting approximating decisions are *statistically consistent*, i.e., they almost surely converge to an optimal solution of the true problem as the number of samples approaches infinity. If, on the other hand, the decisions fail to converge, or converge to a suboptimal solution, then such approximations are unlikely to be meaningful and perform well in practice.

We take a general approach to our analysis, assuming weakly converging sequences of distributional approximations (encompassing SAA), while simultaneously permitting the approximation of other structural aspects of the problem. Due to their recursively defined nature, the approximating SDP equations involve a dependence between an objective function and a *varying* random function. Compared to optimizing the expectation of a *fixed* random function — where consistency follows from suitably uniform versions of the strong law of large numbers [74, 7] — establishing consistency in this setting is more challenging.

This chapter is based on the joint work:

Epi-Consistent Approximation of Stochastic Dynamic Programs

Dominic S. T. Keehan & Johannes O. Royset [71].

Stochastic optimal control problems are faithfully expressed through the solution of SDP equations when the principle of optimality holds [82]. Establishing the consistency of an approximation to SDP equations then entails: (i) showing that a solution to the true equations exists and satisfies the principle of optimality, and (ii) showing that this solution is approached by the solutions of the approximating equations. Typically, (i) is addressed with boundedness assumptions (see, e.g., [18, Chapters 3 and 4] or [113, Chapter 9]), but more generally with monotone convergence [18, Chapter 5] or inf-compactness [48, 49]. In this chapter we address (ii) and provide results which ensure that the solution functions of the approximating SDP equations epi-converge to solution functions of the true SDP equations, and, as a consequence, that approximating control policies are consistent.

There are existing works on the consistent approximation of multistage stochastic optimization problems. For finite-horizon problems, [88, 89] establishes consistency by placing conditions on the scenario trees of the approximating control problem and requiring lower compactness of the objective function over the decision space. These conditions can be quite restrictive; the results may not hold when the support of the underlying distribution has holes [89, Section 5.1], and lower compactness does not hold when the objective function cannot be lower bounded by a linear function which excludes some problems involving saddle functions [66, Theorem 5]. In related work, [103] considers a problem involving consistency in terms of an iterated (rather than joint) limit in the number of random samples in each stage. From a practical perspective this is unnatural as it requires a discretisation where the number of samples in future stages must increase at a faster rate than the number of samples in prior stages, which may be difficult to implement. For infinite-horizon problems, [109] utilises the contraction-mapping property of Bellman operators defined on a complete metric space of bounded functions under the sup-norm. This approach fails when the stage-cost functions are unbounded. We avoid all of these limiting assumptions.

The results of this chapter provide theoretical justification for algorithms which solve approximations of stochastic dynamic programs — we *do not* deal with the practicalities of actually solving these problems computationally. However, stochastic dual dynamic programming–

type algorithms have proven useful for this purpose, especially when the number of stages is not too large [90]. Infinite-horizon problems may also be approached by additionally approximating the number of stages [78] or with specialised algorithms; see, e.g., [85, 109, 37, 68]. Although it is well known that both the algorithmic and statistical rate of convergence in long-horizon problems can be very slow, algorithms may be improved with specific bounding functions [107] and statistical errors follow central limit–like behaviour for certain problem classes [106]. Therefore, it is clear that at least some SDP problems can be feasibly solved to an acceptable accuracy.

The remainder of this chapter is organised as follows. In Section 2.1 we detail mathematical preliminaries and the necessary background on epi-convergence. Section 2.2 considers finite-horizon SDP, presents intermediate results on the epi-convergence of varying sums and expectation functions, and provides conditions under which the solution functions of approximating SDP equations epi-converge to solution functions of the true equations. Section 2.3 considers infinite-horizon SDP and adapts the results of the previous section to this setting using the metric structure imposed by the Attouch–Wets distance. We conclude in Section 2.4 by considering a number of examples, including problems with unbounded stage-cost functions and simultaneously approximated constraints.

2.1 Preliminaries

Let $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{-\infty, \infty\}$. For a closed subset $\mathcal{Z} \subseteq \mathbb{R}^r$, let $\operatorname{fcns}(\mathcal{Z})$ denote the set of $\overline{\mathbb{R}}$-valued functions on $\mathcal{Z}$ excluding $f = \infty$. A function $f \in \operatorname{fcns}(\mathcal{Z})$ is *proper* if $f(z) > -\infty$ for all $z \in \mathcal{Z}$. A sequence of functions $\{f_\nu \in \operatorname{fcns}(\mathcal{Z})\}_{\nu\in\mathbb{N}}$ *epi-converges* to a function $f \in \operatorname{fcns}(\mathcal{Z})$, denoted $f_\nu \xrightarrow{\text{e}} f$, if for each $z \in \mathcal{Z}$ the following are satisfied:

(i) for all $z_\nu \in \mathcal{Z} \to z$, we have $\operatorname{liminf} f_\nu(z_\nu) \geq f(z)$,

(ii) there exists $z_\nu \in \mathcal{Z} \to z$ such that $\operatorname{limsup} f_\nu(z_\nu) \leq f(z)$;

these are the *liminf* and *limsup* conditions of epi-convergence, respectively. The *outer limit* of a sequence of sets, denoted LimOut, is the collection of cluster points to which subsequences of elements within the sets converge. If $f_\nu \xrightarrow{\text{e}} f$, then $\operatorname{LimOut}(\operatorname{argmin} f_\nu) \subseteq \operatorname{argmin} f$, provided that the limit function f is proper. In addition, if $\bigcup_{\nu\in\mathbb{N}}\{z \in \mathcal{Z} : f_\nu(z) < \infty\}$ is bounded, then $\inf f_\nu \to \inf f$ as well; see, e.g., [96, Section 4.E] and a slight extension of [96, Theorem 5.5].

A function $f \in \operatorname{fcns}(\mathcal{Z})$ is *lower semicontinuous* (lsc) if $\operatorname{liminf} f(z_\nu) \geq f(z)$ for every $z_\nu \in \mathcal{Z} \to z$; let $\operatorname{lsc-fcns}(\mathcal{Z})$ denote the subset of $\operatorname{fcns}(\mathcal{Z})$ consisting of lsc functions. In this chapter

we will work with Euclidean balls; for $\bar{z} \in \mathcal{Z}$ and $\delta \geq 0$, let $\mathbb{B}_2(\bar{z}; \delta) \coloneqq \{z \in \mathcal{Z} : \|z - \bar{z}\|_2 \leq \delta\}$. A sequence of functions $\{f_\nu \in \text{lsc-fcns}(\mathcal{Z})\}_{\nu \in \mathbb{N}}$ is *equi-lsc at* $\bar{z} \in \mathcal{Z}$ if either $\operatorname{limsup} f_\nu(\bar{z}) \to -\infty$, or for each $\rho, \varepsilon \in (0, \infty)$, there exists $\bar{\nu} \in \mathbb{N}$ and $\delta > 0$ such that

$$\inf_{z \in \mathbb{B}_2(\bar{z}; \delta)} f_\nu(z) \geq \min\{f_\nu(\bar{z}) - \varepsilon, \rho\} \quad \text{for all } \nu \geq \bar{\nu}.$$

We say that $\{f_\nu\}_{\nu \in \mathbb{N}}$ is *equi-lsc* if it is equi-lsc at every $\bar{z} \in \mathcal{Z}$.

A sequence of functions $\{f_\nu \in \text{fcns}(\mathcal{Z})\}_{\nu \in \mathbb{N}}$ *converges pointwise* to a function $f \in \text{fcns}(\mathcal{Z})$, denoted $f_\nu \xrightarrow{\text{p}} f$, if $f_\nu(z) \to f(z)$ for each $z \in \mathcal{Z}$. If $\{f_\nu\}_{\nu \in \mathbb{N}}$ is equi-lsc, then $f_\nu \xrightarrow{\text{p}} f$ if and only if $f_\nu \xrightarrow{\text{e}} f$ [94, Theorem 7.10]. We will also need the notions of hypo-convergence ($\xrightarrow{\text{h}}$), upper semicontinuity (usc), and equi-upper semicontinuity (equi-usc); these have parallel definitions to those of epi-convergence, lower semicontinuity, and equi-lower semicontinuity, respectively; see [94] for details.

Recall that $\mathfrak{P}(\Xi)$ denotes the set of Borel probability measures on Ξ. We denote the weak convergence of a sequence of distributions $\{\mathbb{P}_\nu \in \mathfrak{P}(\Xi)\}_{\nu \in \mathbb{N}}$ to a distribution $\mathbb{P} \in \mathfrak{P}(\Xi)$ by $\mathbb{P}_\nu \Rightarrow \mathbb{P}$; see [20] and [75, Chapter 13] for details on weak convergence.

2.2 Finite-Horizon Stochastic Dynamic Programming

In this section we consider the finite-horizon version of the general stochastic optimal control problem (SOC). Differing from the notation used throughout the rest of this thesis, in this chapter we use a superscript t to denote indexing with respect to time. Recall that a problem closely related to (SOC) is that of finding a family of real-valued solution functions $V^t : \mathcal{X} \to \mathbb{R}$, $t \in [T]$, to the SDP functional equations

$$V^t(x) = \mathbb{E}_{\mathbb{P}}\Big[\inf_{y \in \mathcal{Y}(x, \boldsymbol{\xi})} \Big\{\varphi(x, y, \boldsymbol{\xi}) + \beta V^{t+1}(y)\Big\}\Big], \quad \text{for all } x \in \mathcal{X},\ t \in [T], \tag{2.1}$$

where $V^{T+1} \coloneqq 0$. (Note that this formulation is equivalent to that of (1.1), except that the function V^t is sought directly, rather than implicitly through the expectation function $\mathbb{E}_{\mathbb{P}}[v^t(\,\cdot\,, \boldsymbol{\xi})]$; this slightly simplifies our analyses.) Given such V^t, $t \in [T]$, a function $y^t : \mathcal{X} \times \Xi \to \mathcal{X}$ with

$$y^t(x, \xi) \in \operatorname*{argmin}_{y \in \mathcal{Y}(x, \xi)} \Big\{\varphi(x, y, \xi) + \beta V^{t+1}(y)\Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi,$$

and with $\xi \mapsto y^t(x,\xi)$ measurable for all $x \in \mathcal{X}$, defines a decision rule at stage t. Repeating this for each stage defines an admissible policy for (SOC). If V^t, $t \in [T]$, satisfies the principle of optimality, then such a policy solves (SOC) and the value $V^1(x^1)$ is equal to its optimal objective value; see, e.g., [113, Theorem 9.2].

For a sequence of probability distributions $\{\mathbb{P}_\nu \in \mathfrak{P}(\Xi)\}_{\nu\in\mathbb{N}}$ approximating $\mathbb{P}$ and a sequence of functions $\{\varphi_\nu : \mathcal{X}\times\mathcal{X}\times\Xi \to \mathbb{R}\}_{\nu\in\mathbb{N}}$ approximating φ, solutions to (2.1) can be approximated by finding a family of real-valued solution functions $V_\nu^t : \mathcal{X} \to \mathbb{R}$, $t \in [T]$, to the approximating SDP functional equations

$$V_\nu^t(x) = \mathbb{E}_{\mathbb{P}_\nu}\Big[\inf_{y\in\mathcal{Y}(x,\boldsymbol{\xi})}\Big\{\varphi_\nu(x,y,\boldsymbol{\xi}) + \beta V_\nu^{t+1}(y)\Big\}\Big], \quad \text{for all } x \in \mathcal{X},\ t \in [T], \tag{2.2}$$

where $V_\nu^{T+1} \coloneqq 0$. Given such V_ν^t, $t \in [T]$, an approximating decision rule $y_\nu^t : \mathcal{X}\times\Xi \to \mathcal{X}$ at stage t is defined via

$$y_\nu^t(x,\xi) \in \operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi_\nu(x,y,\xi) + \beta V_\nu^{t+1}(y)\Big\}, \quad \text{for all } (x,\xi) \in \mathcal{X}\times\Xi,$$

and this therefore defines an approximating policy. Approximating the set-valued mapping $\mathcal{Y}$ can also be treated in the present setting; see Subsection 2.4.3.

The equations in (2.1) feature a *Bellman operator* $B : \text{lsc-fcns}(\mathcal{X}) \to \text{fcns}(\mathcal{X})$ which for $f \in \text{lsc-fcns}(\mathcal{X})$ at $x \in \mathcal{X}$ has value

$$B(f)(x) \coloneqq \mathbb{E}_{\mathbb{P}}\Big[\inf_{y\in\mathcal{Y}(x,\boldsymbol{\xi})}\Big\{\varphi(x,y,\boldsymbol{\xi}) + \beta f(y)\Big\}\Big]; \tag{2.3}$$

using B they may be restated as $V^t = B(V^{t+1})$, $t \in [T]$. It will be helpful to also consider the function $b : \text{lsc-fcns}(\mathcal{X}) \to \text{fcns}(\mathcal{X}\times\Xi)$ which for $f \in \text{lsc-fcns}(\mathcal{X})$ at $(x,\xi) \in \mathcal{X}\times\Xi$ has value

$$b(f)(x,\xi) \coloneqq \inf_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi(x,y,\xi) + \beta f(y)\Big\}.$$

By definition, $B(f) = \mathbb{E}_{\mathbb{P}}[b(f)(\,\cdot\,,\boldsymbol{\xi})]$. We similarly define $B_\nu : \text{lsc-fcns}(\mathcal{X}) \to \text{fcns}(\mathcal{X})$, $\nu \in \mathbb{N}$, and $b_\nu : \text{lsc-fcns}(\mathcal{X}) \to \text{fcns}(\mathcal{X}\times\Xi)$, $\nu \in \mathbb{N}$, for the approximating equations in (2.2).

To ensure that the solution functions to (2.1) and (2.2) are well defined, we make the following assumption.

Assumption 2.1. *There exists a set* $\mathcal{V} \subseteq \text{lsc-fcns}(\mathcal{X})$ *with* $0 \in \mathcal{V}$ *satisfying the following condition:*

> *For each* $V \in \mathcal{V}$ *and* $x \in \mathcal{X}$ *the function* $b(V)(x, \cdot)$ *is integrable with respect to* $\mathbb{P}$*, and the Bellman operator* B *maps* $\mathcal{V}$ *into itself.*

Moreover, for each $\nu \in \mathbb{N}$*, the set* $\mathcal{V}$ *satisfies the equivalent condition for* b_ν*,* $\mathbb{P}_\nu$*, and* B_ν*.*

Under Assumption 2.1 the function $V^t = B(V^{t+1}) = \mathbb{E}_{\mathbb{P}}[b(V^{t+1})(\,\cdot\,, \boldsymbol{\xi})] \in \mathcal{V}$ is real valued for each $V^{t+1} \in \mathcal{V}$. It follows via induction that the solution functions to (2.1) are well defined. Similar observations apply to (2.2). Section 2.4 furnishes examples where the assumption holds.

We seek to characterise sequences of approximating families of real-valued solution functions $\{V_\nu^t,\, t \in [T]\}_{\nu \in \mathbb{N}}$ to (2.2), consistent in the sense that the inclusion

$$\text{LimOut}\Big(\operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)}\Big\{\varphi_\nu(x,y,\xi) + \beta V_\nu^{t+1}(y)\Big\}\Big) \subseteq \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)}\Big\{\varphi(x,y,\xi) + \beta V^{t+1}(y)\Big\}, \quad \text{for all } (x,\xi) \in \mathcal{X} \times \Xi,$$

holds for some family of real-valued solution functions V^t, $t \in [T]$, to (2.1). In other words, each convergent sequence of decisions obtained from the approximating problems will converge to an optimal decision for the true SDP equations. Whether or not solutions to (2.2) satisfy the principle of optimality relative to an underlying approximation of (SOC) is not strictly required — the stochastic optimal control problem (SOC) need not even be well defined. In this regard, throughout this chapter we only assume that $\mathcal{X}$ and Ξ are closed, and that $\mathcal{Y}$ is nonempty and compact valued. However, in the examples of Section 2.4 we return to the question of the principle of optimality.

The equations in (2.2) involve a sum of two functions that both vary with ν. To ensure that epi-convergence is preserved under this operation, we prove the following lemma. Its conclusion is standard, although its proof is slightly nonstandard in that it applies to functions with domains that are strict subsets of finite-dimensional Euclidean spaces.

Lemma 2.1. *For functions* $f, f_\nu, g, g_\nu : \mathcal{X} \to \mathbb{R}$*,* $\nu \in \mathbb{N}$*, if* $f_\nu \xrightarrow{\text{e}} f$*,* $f_\nu \xrightarrow{\text{p}} f$*,* $g_\nu \xrightarrow{\text{e}} g$*, and* $g_\nu \xrightarrow{\text{p}} g$*, then*

$$\inf_{y \in \mathcal{Y}(x,\xi)}\Big\{f_\nu(y) + \beta g_\nu(y)\Big\} \to \inf_{y \in \mathcal{Y}(x,\xi)}\Big\{f(y) + \beta g(y)\Big\}, \quad \textit{for all } (x,\xi) \in \mathcal{X} \times \Xi.$$

Proof. Fix $(x,\xi) \in \mathcal{X} \times \Xi$. For each $\nu \in \mathbb{N}$, define $\psi_\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ by $\psi_\nu(y) \coloneqq f_\nu(y) + \beta g_\nu(y)$ if $y \in \mathcal{Y}(x,\xi)$, and by $\psi_\nu(y) \coloneqq \infty$ otherwise. Define $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ similarly. Following [96,

Proposition 4.19(a)], we first show that $\psi_\nu \xrightarrow{\text{e}} \psi$. For $y \in \mathbb{R}^n$, if $y_\nu \in \mathbb{R}^n \to y$, then either $\operatorname{liminf} \psi_\nu(y_\nu) = \infty \geq \psi(y)$ and the liminf condition of epi-convergence is satisfied, or there exists a subsequence with $y_{\nu_\kappa} \in \mathcal{Y}(x,\xi) \to y$ such that $\operatorname{liminf} \psi_\nu(y_\nu) = \operatorname{liminf} \psi_{\nu_\kappa}(y_{\nu_\kappa})$. In the second case we construct a padded version of the sequence of indices $\{\nu_\kappa\}_{\kappa\in\mathbb{N}}$ as follows: for each $\kappa \in \mathbb{N}$, set $\nu_{\kappa_\ell} = \min_{i\in\mathbb{N}}\{\nu_i : \nu_i \geq \kappa\}$. Now, $\{(\nu_\kappa, \nu_\kappa)\}_{\kappa\in\mathbb{N}}$ is a subsequence of $\{(\kappa, \nu_{\kappa_\ell})\}_{\kappa\in\mathbb{N}}$, which implies that

$$\operatorname{liminf} f_{\nu_\kappa}(y_{\nu_\kappa}) \geq \operatorname{liminf} f_\kappa(y_{\nu_{\kappa_\ell}}) \quad \text{and} \quad \operatorname{liminf} \beta g_{\nu_\kappa}(y_{\nu_\kappa}) \geq \operatorname{liminf} \beta g_\kappa(y_{\nu_{\kappa_\ell}}).$$

Clearly, $y_{\nu_{\kappa_\ell}} \in \mathcal{Y}(x,\xi) \to y$ as well, and since $f_\nu \xrightarrow{\text{e}} f$ and $g_\nu \xrightarrow{\text{e}} g$, we have

$$\operatorname{liminf} f_\kappa(y_{\nu_{\kappa_\ell}}) \geq f(y) \quad \text{and} \quad \operatorname{liminf} \beta g_\kappa(y_{\nu_{\kappa_\ell}}) \geq \beta g(y).$$

Together these yield

$$\begin{aligned}
\operatorname{liminf} \psi_\nu(y_\nu) &= \operatorname{liminf}\big(f_{\nu_\kappa}(y_{\nu_\kappa}) + \beta g_{\nu_\kappa}(y_{\nu_\kappa})\big) \\
&\geq \operatorname{liminf} f_{\nu_\kappa}(y_{\nu_\kappa}) + \operatorname{liminf} \beta g_{\nu_\kappa}(y_{\nu_\kappa}) \\
&\geq \operatorname{liminf} f_\kappa(y_{\nu_{\kappa_\ell}}) + \operatorname{liminf} \beta g_\kappa(y_{\nu_{\kappa_\ell}}) \\
&\geq f(y) + \beta g(y) \\
&= \psi(y).
\end{aligned}$$

The limsup condition of epi-convergence follows for each $y \in \mathbb{R}^n$ when taking the sequence with $y_\nu = y$ for each $\nu \in \mathbb{N}$; either $y \in \mathcal{Y}(x,\xi)$ and

$$\begin{aligned}
\operatorname{limsup} \psi_\nu(y_\nu) &= \operatorname{limsup}\big(f_\nu(y) + \beta g_\nu(y)\big) \\
&\leq \operatorname{limsup} f_\nu(y) + \operatorname{limsup} \beta g_\nu(y) \\
&= f(y) + \beta g(y) \\
&= \psi(y),
\end{aligned}$$

where the second equality holds since $f_\nu \xrightarrow{\text{p}} f$ and $g_\nu \xrightarrow{\text{p}} g$; or $y \notin \mathcal{Y}(x,\xi)$ and

$$\operatorname{limsup} \psi_\nu(y_\nu) = \operatorname{limsup} \psi_\nu(y) = \infty = \psi(y).$$

We now establish the claimed convergence. The function ψ is proper since it takes finite values on the nonempty set $\mathcal{Y}(x,\xi)$ and ∞ elsewhere. Moreover, $\{y \in \mathbb{R}^n : \psi_\nu(y) < \infty\} = \mathcal{Y}(x,\xi)$ for each $\nu \in \mathbb{N}$, and so $\bigcup_{\nu\in\mathbb{N}}\{y \in \mathbb{R}^n : \psi_\nu(y) < \infty\} = \mathcal{Y}(x,\xi)$. As assumed throughout, this set is compact and therefore bounded. Thus, [96, Theorem 5.5(d)] applies to $\{\psi, \psi_\nu;\, \nu \in \mathbb{N}\}$ and

$$\inf_{y\in\mathcal{Y}(x,\xi)}\Big\{f_\nu(y) + \beta g_\nu(y)\Big\} = \inf \psi_\nu \to \inf \psi = \inf_{y\in\mathcal{Y}(x,\xi)}\Big\{f(y) + \beta g(y)\Big\},$$

which proves the result. □

The solutions to the equations in (2.2) are expectation functions with a probability distribution and an integrand that vary with ν. We use the following result of Feinberg, Kasyanov, & Royset [47] to establish epi-convergence for functions of this type. In our setting $\mathcal{X} \subseteq \mathbb{R}^n$ and $\Xi \subseteq \mathbb{R}^m$ are the closed state and outcome spaces.

Proposition 2.1 ([47], Theorem 3.7)**.** *For probability distributions $\mathbb{P}, \mathbb{P}_\nu \in \mathfrak{P}(\Xi)$ and functions $f, f_\nu : \mathcal{X} \times \Xi \to \overline{\mathbb{R}}$, $\nu \in \mathbb{N}$, suppose that $\mathbb{P}_\nu \Rightarrow \mathbb{P}$ and the following conditions are satisfied for each $\bar{x} \in \mathcal{X}$:*

(i) *The functions $f(\bar{x}, \cdot)$, $f_\nu(\bar{x}, \cdot)$, $\nu \in \mathbb{N}$ are measurable.*

(ii) *We have*

$$\liminf_{\tau\to-\infty} \liminf_{(\nu,x)\to(\infty,\bar{x})} \mathbb{E}_{\mathbb{P}_\nu}\Big[f_\nu(x,\boldsymbol{\xi}) \cdot \mathbb{1}\{f_\nu(x,\boldsymbol{\xi}) \le \tau\}\Big] = 0 \tag{2.4}$$

and for $\mathbb{P}$-almost every $\bar{\xi} \in \Xi$,

$$\liminf_{(\nu,x,\xi)\to(\infty,\bar{x},\bar{\xi})} f_\nu(x,\xi) \ge f(\bar{x},\bar{\xi}). \tag{2.5}$$

(iii) *There exists a sequence $x_\nu \in \mathcal{X} \to \bar{x}$ such that*

$$\limsup_{\tau\to\infty} \limsup_{\nu\to\infty} \mathbb{E}_{\mathbb{P}_\nu}\Big[f_\nu(x_\nu,\boldsymbol{\xi}) \cdot \mathbb{1}\{f_\nu(x_\nu,\boldsymbol{\xi}) \ge \tau\}\Big] = 0 \tag{2.6}$$

and for $\mathbb{P}$-almost every $\bar{\xi} \in \Xi$,

$$\limsup_{(\nu,\xi)\to(\infty,\bar{\xi})} f_\nu(x_\nu,\xi) \le f(\bar{x},\bar{\xi}). \tag{2.7}$$

Then

$$\mathbb{E}_{\mathbb{P}_\nu}[f_\nu(\,\cdot\,,\boldsymbol{\xi})] \xrightarrow{\text{e}} \mathbb{E}_{\mathbb{P}}[f(\,\cdot\,,\boldsymbol{\xi})].$$

We are now ready to prove our first main result.

Theorem 2.1. *Let Assumption 2.1 hold and $\mathbb{P}_\nu \Rightarrow \mathbb{P}$, $\varphi_\nu \xrightarrow{\mathrm{e}} \varphi$, and $\varphi_\nu \xrightarrow{\mathrm{p}} \varphi$. Suppose that for each stage $t \in [T]$ the sequence of real-valued solution functions $\{V_\nu^{t+1}\}_{\nu\in\mathbb{N}}$ to (2.2) is equi-lsc and satisfies the following conditions for each $\bar{x} \in \mathcal{X}$:*

(i) *We have*

$$\liminf_{\tau\to-\infty} \liminf_{(\nu,x)\to(\infty,\bar{x})} \mathbb{E}_{\mathbb{P}_\nu}\Big[b_\nu(V_\nu^{t+1})(x,\boldsymbol{\xi}) \cdot \mathbb{1}\{b_\nu(V_\nu^{t+1})(x,\boldsymbol{\xi}) \le \tau\}\Big] = 0, \tag{2.8}$$

and that the sequence $\{b_\nu(V_\nu^{t+1})\}_{\nu\in\mathbb{N}}$ of lsc functions is equi-lsc.

(ii) *We have*

$$\limsup_{\tau\to\infty} \limsup_{\nu\to\infty} \mathbb{E}_{\mathbb{P}_\nu}\Big[b_\nu(V_\nu^{t+1})(\bar{x},\boldsymbol{\xi}) \cdot \mathbb{1}\{b_\nu(V_\nu^{t+1})(\bar{x},\boldsymbol{\xi}) \ge \tau\}\Big] = 0, \tag{2.9}$$

and that the sequence $\{b_\nu(V_\nu^{t+1})(\bar{x},\,\cdot\,)\}_{\nu\in\mathbb{N}}$ of usc functions is equi-usc.

Then, for each time $t \in [T]$, the function $V_\nu^{t+1} \xrightarrow{\mathrm{e}} V^{t+1}$, and

$$\mathrm{LimOut}\Big(\operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi_\nu(x,y,\xi) + \beta V_\nu^{t+1}(y)\Big\}\Big) \subseteq \operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi(x,y,\xi) + \beta V^{t+1}(y)\Big\} \neq \varnothing$$

for all $(x,\xi) \in \mathcal{X}\times\Xi$.

Proof. We proceed by induction in t on the claim $V_\nu^{t+1} \xrightarrow{\mathrm{e}} V^{t+1}$. We have $V_\nu^{T+1} = 0 \xrightarrow{\mathrm{e}} 0 = V^{T+1}$ which is the base case. Assume that $V_\nu^{t+1} \xrightarrow{\mathrm{e}} V^{t+1}$ for some $t \in [T]$. With $\{V_\nu^{t+1}\}_{\nu\in\mathbb{N}}$ equi-lsc, it follows that $V_\nu^{t+1} \xrightarrow{\mathrm{p}} V^{t+1}$ as well [94, Theorem 7.10]. Since $\varphi_\nu \xrightarrow{\mathrm{e}} \varphi$ and $\varphi_\nu \xrightarrow{\mathrm{p}} \varphi$, for every $(x,\xi) \in \mathcal{X}\times\Xi$ it holds that $\varphi_\nu(x,\,\cdot\,,\xi) \xrightarrow{\mathrm{e}} \varphi(x,\,\cdot\,,\xi)$ and $\varphi_\nu(x,\,\cdot\,,\xi) \xrightarrow{\mathrm{p}} \varphi(x,\,\cdot\,,\xi)$, so Lemma 2.1 applies to

$$\begin{aligned} b_\nu(V_\nu^{t+1})(x,\xi) &= \inf_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi_\nu(x,y,\xi) + \beta V_\nu^{t+1}(y)\Big\} \\ &\to \inf_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi(x,y,\xi) + \beta V^{t+1}(y)\Big\} = b(V^{t+1})(x,\xi), \end{aligned}$$

and we deduce that $b_\nu(V_\nu^{t+1}) \xrightarrow{\mathrm{p}} b(V^{t+1})$. By (i) in the statement of the theorem, the sequence $\{b_\nu(V_\nu^{t+1})\}_{\nu\in\mathbb{N}}$ is equi-lsc and it follows that $b_\nu(V_\nu^{t+1}) \xrightarrow{\mathrm{e}} b(V^{t+1})$ [94, Theorem 7.10]. To conclude that $V_\nu^t \xrightarrow{\mathrm{e}} V^t$, we apply Proposition 2.1 to the distributions $\mathbb{P}, \mathbb{P}_\nu$ and functions $b(V^{t+1}), b_\nu(V_\nu^{t+1})$, $\nu \in \mathbb{N}$, checking the conditions (i)–(iii) as follows.

(i): Satisfied due to the integrability (and therefore measurability) provided by Assumption 2.1. (ii): The asymptotic inf-integrability statement (2.4) is satisfied due to (i) in the statement of the theorem, and the liminf statement (2.5) is satisfied since $b_\nu(V_\nu^{t+1}) \xrightarrow{\text{e}} b(V^{t+1})$ implies, for every $(\bar{x}, \bar{\xi}) \in \mathcal{X} \times \Xi$, that

$$\begin{aligned} &\text{for all } (x_\nu, \xi_\nu) \to (\bar{x}, \bar{\xi}), \quad \operatorname{liminf} b_\nu(V_\nu^{t+1})(x_\nu, \xi_\nu) \geq b(V^{t+1})(\bar{x}, \bar{\xi}) \\ &\iff \operatorname*{liminf}_{(\nu, x, \xi) \to (\infty, \bar{x}, \bar{\xi})} b_\nu(V_\nu^{t+1})(x, \xi) \geq b(V^{t+1})(\bar{x}, \bar{\xi}). \end{aligned}$$

(iii): For each $\bar{x} \in \mathcal{X}$, take the sequence with $x_\nu = \bar{x}$ for each $\nu \in \mathbb{N}$. Due to (ii) in the statement of the theorem, the asymptotic sup-integrability statement (2.6) is satisfied. With $b_\nu(V_\nu^{t+1}) \xrightarrow{\text{p}} b(V^{t+1})$, we have that $b_\nu(V_\nu^{t+1})(\bar{x}, \cdot\,) \xrightarrow{\text{p}} b(V^{t+1})(\bar{x}, \cdot\,)$, and again by (ii), the sequence $\{b_\nu(V_\nu^{t+1})(\bar{x}, \cdot\,)\}_{\nu \in \mathbb{N}}$ is equi-usc. Whence, $b_\nu(V_\nu^{t+1})(\bar{x}, \cdot\,) \xrightarrow{\text{h}} b(V^{t+1})(\bar{x}, \cdot\,)$ [94, Theorem 7.10]. This implies, for every $\bar{\xi} \in \Xi$, that

$$\begin{aligned} &\text{for all } \xi_\nu \to \bar{\xi}, \quad \operatorname{limsup} b_\nu(V_\nu^{t+1})(\bar{x}, \xi_\nu) \leq b(V^{t+1})(\bar{x}, \bar{\xi}) \\ &\iff \operatorname*{limsup}_{(\nu, \xi) \to (\infty, \bar{\xi})} b_\nu(V_\nu^{t+1})(\bar{x}, \xi) \leq b(V^{t+1})(\bar{x}, \bar{\xi}), \end{aligned}$$

which is the limsup statement (2.7). Hence, Proposition 2.1 applies, and we conclude that

$$V_\nu^t = B_\nu(V_\nu^{t+1}) = \mathbb{E}_{\mathbb{P}_\nu}[b_\nu(V_\nu^{t+1})(\,\cdot\,, \boldsymbol{\xi})] \xrightarrow{\text{e}} \mathbb{E}_{\mathbb{P}}[b(V^{t+1})(\,\cdot\,, \boldsymbol{\xi})] = B(V^{t+1}) = V^t.$$

Reasoning similar to that in the proof of Lemma 2.1 as well as [96, Theorem 5.5(b)] then shows that the inclusion

$$\operatorname{LimOut}\Big(\operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \big\{\varphi_\nu(x, y, \xi) + \beta V_\nu^{t+1}(y)\big\}\Big) \subseteq \operatorname*{argmin}_{y \in \mathcal{Y}(x,\xi)} \big\{\varphi(x, y, \xi) + \beta V^{t+1}(y)\big\} \neq \varnothing,$$
$$\text{for all } (x, \xi) \in \mathcal{X} \times \Xi,$$

holds. To see that this argmin is nonempty, observe that $\varphi(x, \cdot\,, \xi) + \beta V^{t+1}(\,\cdot\,)$ is lsc and real valued, and that $\mathcal{Y}(x, \xi)$ is nonempty and compact by assumption. This proves the result. □

The statements (2.8) and (2.9) in the conditions (i) and (ii) of Theorem 2.1 are asymptotic semi-integrability statements that hold, for instance, under asymptotic local boundedness. Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots \sim \mathbb{P}$ be independent and identically distributed random vectors and suppose that

$\{\xi_1, \xi_2, \ldots\}$ is an outcome of this random sequence. Set $\mathbb{P}_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} \mathbb{1}_{\xi_i}$ for each $\nu \in \mathbb{N}$, so that $\mathbb{P}_\nu$ is the ν-th empirical distribution (as in SAA). Note that $\mathbb{P}_\nu \Rightarrow \mathbb{P}$ for almost every outcome $\{\xi_1, \xi_2, \ldots\}$; see, e.g., [20, Problem 3.1]. For a given $\bar{x} \in \mathcal{X}$, assume that there exists an integrable function $f : \Xi \to \mathbb{R}$ and a constant $\delta > 0$ such that $f(\xi) \leq b_\nu(V_\nu^{t+1})(x, \xi)$ holds for all $(x, \xi) \in (\mathbb{B}_2(\bar{x}; \delta) \cap \mathcal{X}) \times \Xi$ and $\nu \in \mathbb{N}$ sufficiently large. Then

$$\begin{aligned}
0 &\geq \liminf_{\tau \to -\infty} \liminf_{(\nu,x) \to (\infty,\bar{x})} \mathbb{E}_{\mathbb{P}_\nu}\Big[b_\nu(V_\nu^{t+1})(x, \boldsymbol{\xi}) \cdot \mathbb{1}\big\{b_\nu(V_\nu^{t+1})(x, \boldsymbol{\xi}) \leq \tau\big\}\Big] \\
&\geq \liminf_{\tau \to -\infty} \liminf_{\nu \to \infty} \mathbb{E}_{\mathbb{P}_\nu}\Big[f(\boldsymbol{\xi}) \cdot \mathbb{1}\big\{f(\boldsymbol{\xi}) \leq \tau\big\}\Big] \\
&= \liminf_{\tau \to -\infty} \mathbb{E}_{\mathbb{P}}\Big[f(\boldsymbol{\xi}) \cdot \mathbb{1}\big\{f(\boldsymbol{\xi}) \leq \tau\big\}\Big] \\
&= 0,
\end{aligned}$$

where the first equality follows for almost every outcome $\{\xi_1, \xi_2, \ldots\}$ due to the strong law of large numbers, and the second equality holds since $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$ is finite. Moreover, the equi-semicontinuity conditions of Theorem 2.1 can often be shown to hold using arguments involving convexity and concavity; see, e.g., [97]. In particular, it is sufficient for the functions to be locally Lipschitz continuous with common moduli. More examples follow in Section 2.4.

2.3 Infinite-Horizon Stochastic Dynamic Programming

In this section we consider the infinite-horizon version of the general stochastic optimal control problem (SOC). A closely related problem is again that of finding a real-valued solution function $V : \mathcal{X} \to \mathbb{R}$ to the SDP functional equation

$$V(x) = \mathbb{E}_{\mathbb{P}}\Big[\inf_{y \in \mathcal{Y}(x, \boldsymbol{\xi})} \Big\{\varphi(x, y, \boldsymbol{\xi}) + \beta V(y)\Big\}\Big], \quad \text{for all } x \in \mathcal{X}. \tag{2.10}$$

Solutions to (2.10) can be approximated by finding a real-valued function $V_\nu : \mathcal{X} \to \mathbb{R}$ satisfying the approximating SDP equation

$$V_\nu(x) = \mathbb{E}_{\mathbb{P}_\nu}\Big[\inf_{y \in \mathcal{Y}(x, \boldsymbol{\xi})} \Big\{\varphi_\nu(x, y, \boldsymbol{\xi}) + \beta V_\nu(y)\Big\}\Big], \quad \text{for all } x \in \mathcal{X}. \tag{2.11}$$

In contrast to the finite-horizon SDP equations, (2.10) and (2.11) are functional fixed-point equations in their associated Bellman operators; they may be restated as $V = B(V)$ and $V_\nu = B_\nu(V_\nu)$, respectively. For a set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathcal{X})$, even if $B(V)$ is a well-defined function for

each $V \in \mathcal{V}$, the operator B may not have a fixed point in $\mathcal{V}$. For now, given a set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathcal{X})$, we make the following integrability assumption.

Assumption 2.2. *The set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathcal{X})$ satisfies the following condition:*

> *For each $V \in \mathcal{V}$ and $x \in \mathcal{X}$ the function $b(V)(x, \cdot)$ is integrable with respect to the probability distribution $\mathbb{P}$, and the Bellman operator B maps $\mathcal{V}$ into itself.*

Moreover, for each $\nu \in \mathbb{N}$, the set $\mathcal{V}$ satisfies the equivalent condition for b_ν, $\mathbb{P}_\nu$, and B_ν.

Assumption 2.2 differs from Assumption 2.1 in that the existence of such a set $\mathcal{V}$ is not presupposed. For results on fixed-point problems it is natural to delegate the verification of the existence of fixed points to the user of said results, as the analysis often requires utilising context-dependent properties of the associated fixed-point operators. Additionally, the inclusion $0 \in \mathcal{V}$ is not present as the terminal condition $V^{T+1} = 0$ is no longer relevant in the infinite-horizon setting.

To study the asymptotic behaviour of sequences of solutions to functional fixed-point problems, we metrise epi-convergence. To this end, let $z_{\text{ctr}} \in \mathcal{X} \times \mathbb{R}$, and for a point $\bar{z} \in \mathcal{X} \times \mathbb{R}$ and a set $\mathcal{Z} \subseteq \mathcal{X} \times \mathbb{R}$, let $\operatorname{dist}(\bar{z}, \mathcal{Z}) \coloneqq \inf_{z \in \mathcal{Z}} \|z - \bar{z}\|_2$. Further, for a function $f \in \text{lsc-fcns}(\mathcal{X})$, denote its *epigraph* by $\operatorname{epi} f \coloneqq \{(x, \tau) \in \mathcal{X} \times \mathbb{R} : f(x) \leq \tau\}$. The *Attouch–Wets distance* [9, 10, 8] between two functions $f, g \in \text{lsc-fcns}(\mathcal{X})$ is then

$$\mathbb{dl}(f, g) \coloneqq \int_0^\infty \max_{z \in \mathbb{B}_2(z_{\text{ctr}}; \rho)} \Big| \operatorname{dist}(z, \operatorname{epi} f) - \operatorname{dist}(z, \operatorname{epi} g) \Big| \cdot \exp(-\rho)\, d\rho,$$

and this defines a metric on lsc-fcns($\mathcal{X}$); for $f, f_\nu \in \text{lsc-fcns}(\mathcal{X})$, $\nu \in \mathbb{N}$, we have $f_\nu \xrightarrow{\text{e}} f$ if and only if $\mathbb{dl}(f_\nu, f) \to 0$. Moreover, under $\mathbb{dl}$, closed and bounded subsets of lsc-fcns($\mathcal{X}$) are compact [94, Proposition 4.45, Theorem 7.58]. Closed subsets include those with elements that are convex, concave, pointwise bounded, or locally Lipschitz–continuous functions; see, e.g., [97, Section 4]. The boundedness of subsets can be established by appealing to the inequality $\mathbb{dl}(f, g) \leq 1 + \max\{\operatorname{dist}(z_{\text{ctr}}, \operatorname{epi} f), \operatorname{dist}(z_{\text{ctr}}, \operatorname{epi} g)\}$ [95, Proposition 3.1]. For instance, if $z_{\text{ctr}} \in \operatorname{epi} f \cap \operatorname{epi} g$, then $\max\{\operatorname{dist}(z_{\text{ctr}}, \operatorname{epi} f), \operatorname{dist}(z_{\text{ctr}}, \operatorname{epi} g)\} = 0$ and $\mathbb{dl}(f, g) \leq 1$, so it suffices to show that all of the epigraphs share a common point. The choice of z_{ctr} alters the numerical value of $\mathbb{dl}$, but not the resulting topology.

For fixed-point functions $V = B(V)$, $V_\nu = B_\nu(V_\nu)$, $\nu \in \mathbb{N}$, we have $V_\nu \xrightarrow{\text{e}} V$ if and only if $B_\nu(V_\nu) \xrightarrow{\text{e}} B(V)$. This motivates the following continuous epi-convergence result in which we utilise the metric properties of $\mathbb{dl}$. Note that the expression $V = B(V)$ is understood to

mean $\mathbb{dl}(V, B(V)) = 0$. Since $\mathbb{dl}$ is a metric over lsc-fcns($\mathcal{X}$), with $V, B(V) \in$ lsc-fcns($\mathcal{X}$), the functions V and $B(V)$ agree pointwise, which is concordant with the definition (2.10).

Lemma 2.2. *Let Assumption 2.2 hold. For a set $\mathcal{V} \subseteq$ lsc-fcns($\mathcal{X}$) that is closed under $\mathbb{dl}$, if $B_\nu(V_\nu) \xrightarrow{\text{e}} B(V)$ whenever $V_\nu \in \mathcal{V} \xrightarrow{\text{e}} V$, then*

$$\operatorname{LimOut}\{V \in \mathcal{V} : V = B_\nu(V)\} \subseteq \{V \in \mathcal{V} : V = B(V)\}.$$

In addition, if $\mathcal{V}$ is compact under $\mathbb{dl}$ and each B_ν has a fixed point $V_\nu = B_\nu(V_\nu)$, then every cluster point of $\{V_\nu\}_{\nu\in\mathbb{N}}$ is contained in $\{V \in \mathcal{V} : V = B(V)\}$ and this set is nonempty.

Proof. We view $\mathcal{V} \subseteq$ lsc-fcns($\mathcal{X}$) as a metric space under $\mathbb{dl}$, and this metrises epi-convergence. Let $V_{\nu_\kappa} \xrightarrow{\text{e}} \bar{V} \in \mathcal{V}$ be a convergent subsequence of elements within the sequence of sets $\{\{V \in \mathcal{V} : V = B_\nu(V)\}\}_{\nu\in\mathbb{N}}$. If no such subsequence exists, then the claim is vacuously true. We construct a padded version of the sequence of indices $\{\nu_\kappa\}_{\kappa\in\mathbb{N}}$ as follows: for each $\kappa \in \mathbb{N}$, set $\nu_{\kappa_\ell} = \min_{i\in\mathbb{N}}\{\nu_i : \nu_i \geq \kappa\}$. Clearly, $V_{\nu_{\kappa_\ell}} \in \mathcal{V} \xrightarrow{\text{e}} \bar{V}$, and it follows that $B_\kappa(V_{\nu_{\kappa_\ell}}) \xrightarrow{\text{e}} B(\bar{V})$. Observe that $\{(\nu_\kappa, \nu_\kappa)\}_{\kappa\in\mathbb{N}}$ is a subsequence of $\{(\kappa, \nu_{\kappa_\ell})\}_{\kappa\in\mathbb{N}}$. But this means that $B_{\nu_\kappa}(V_{\nu_\kappa}) = V_{\nu_\kappa} \xrightarrow{\text{e}} \bar{V}$ has the same limit as $B_\kappa(V_{\nu_{\kappa_\ell}}) \xrightarrow{\text{e}} B(\bar{V})$. Thus, $\bar{V} = B(\bar{V})$, and the cluster point $\bar{V}$ is a fixed point of B. Since the subsequence $\{V_{\nu_\kappa}\}_{\kappa\in\mathbb{N}}$ was arbitrary, we have shown that

$$\operatorname{LimOut}\{V \in \mathcal{V} : V = B_\nu(V)\} \subseteq \{V \in \mathcal{V} : V = B(V)\}.$$

If in addition $\mathcal{V}$ is compact under $\mathbb{dl}$ and each B_ν has a fixed point $V_\nu = B_\nu(V_\nu)$, then a convergent subsequence $V_{\nu_\kappa} \xrightarrow{\text{e}} \bar{V} \in \mathcal{V}$ of $\{V_\nu\}_{\nu\in\mathbb{N}}$ exists. Identical reasoning to that above then shows that the cluster point $\bar{V}$ is a fixed point of B. □

We are now ready to prove our second main result.

Theorem 2.2. *Let Assumption 2.2 hold and $\mathbb{P}_\nu \Rightarrow \mathbb{P}$, $\varphi_\nu \xrightarrow{\text{e}} \varphi$, and $\varphi_\nu \xrightarrow{\text{p}} \varphi$. Suppose that the set $\mathcal{V} \subseteq$ lsc-fcns($\mathcal{X}$) is closed under $\mathbb{dl}$. Further, suppose that every epi-convergent sequence $\{V_\nu \in \mathcal{V}\}_{\nu\in\mathbb{N}}$ is equi-lsc and satisfies the following conditions for each $\bar{x} \in \mathcal{X}$:*

(i) *We have*

$$\liminf_{\tau\to-\infty} \liminf_{(\nu,x)\to(\infty,\bar{x})} \mathbb{E}_{\mathbb{P}_\nu}\Big[b_\nu(V_\nu)(x, \boldsymbol{\xi}) \cdot \mathbb{1}\{b_\nu(V_\nu)(x, \boldsymbol{\xi}) \leq \tau\}\Big] = 0, \tag{2.12}$$

and that the sequence $\{b_\nu(V_\nu)\}_{\nu\in\mathbb{N}}$ of lsc functions is equi-lsc.

(ii) *We have*

$$\limsup_{\tau\to\infty}\limsup_{\nu\to\infty}\mathbb{E}_{\mathbb{P}_\nu}\Big[b_\nu(V_\nu)(\bar{x},\boldsymbol{\xi})\cdot\mathbb{1}\big\{b_\nu(V_\nu)(\bar{x},\boldsymbol{\xi})\geq\tau\big\}\Big]=0, \tag{2.13}$$

and that the sequence $\{b_\nu(V_\nu)(\bar{x},\,\cdot\,)\}_{\nu\in\mathbb{N}}$ *of usc functions is equi-usc.*

Then

$$\operatorname{LimOut}\big\{V\in\mathcal{V}:V=B_\nu(V)\big\}\subseteq\big\{V\in\mathcal{V}:V=B(V)\big\}.$$

In addition, if $\mathcal{V}$ *is compact under* dl *and each* B_ν *has a fixed point* $V_\nu=B_\nu(V_\nu)$*, then every cluster point of* $\{V_\nu\}_{\nu\in\mathbb{N}}$ *is contained in* $\{V\in\mathcal{V}:V=B(V)\}$ *and this set is nonempty.*

Proof. Let $V_\nu\in\mathcal{V}\xrightarrow{\text{e}}V$. Following the proof of Theorem 2.1, we deduce that $b_\nu(V_\nu)\xrightarrow{\text{p}}b(V)$ and $b_\nu(V_\nu)\xrightarrow{\text{e}}b(V)$, and together with (i) and (ii), that Proposition 2.1 applies to $\mathbb{P}$, $\mathbb{P}_\nu$, $\nu\in\mathbb{N}$ and $b(V)$, $b_\nu(V_\nu)$, $\nu\in\mathbb{N}$. Thus,

$$B_\nu(V_\nu)=\mathbb{E}_{\mathbb{P}_\nu}\big[b_\nu(V_\nu)(\,\cdot\,,\boldsymbol{\xi})\big]\xrightarrow{\text{e}}\mathbb{E}_{\mathbb{P}}\big[b(V)(\,\cdot\,,\boldsymbol{\xi})\big]=B(V).$$

Hence, Lemma 2.2 applies, and this proves the result. □

While Theorem 2.2 can be used to conclude that the cluster points of sequences of solution functions to the approximating fixed-point problems (2.11) solve the true fixed-point problem (2.10), making statements about sequences of approximating decision rules obtained from the approximating solution functions is more delicate. We summarise the possibilities in the following proposition.

Proposition 2.2. *Let the conditions in the statement of Theorem 2.2 hold, including* $\mathcal{V}$ *being compact under* dl *and each* B_ν *having a fixed point* $V_\nu=B_\nu(V_\nu)$*. Then the following are true:*

(i) *If* $V_{\nu_\kappa}\xrightarrow{\text{e}}V$ *is an epi-convergent subsequence of* $\{V_\nu=B_\nu(V_\nu)\}_{\nu\in\mathbb{N}}$*, then* $V=B(V)$ *and*

$$\operatorname{LimOut}\Big(\operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi_{\nu_\kappa}(x,y,\xi)+\beta V_{\nu_\kappa}(y)\Big\}\Big)\subseteq\operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\Big\{\varphi(x,y,\xi)+\beta V(y)\Big\}\neq\varnothing,$$

for all $(x,\xi)\in\mathcal{X}\times\Xi$.

(ii) *If the fixed point $V = B(V)$ is unique, then*

$$\operatorname{LimOut}\Big(\operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\big\{\varphi_\nu(x,y,\xi)+\beta V_\nu(y)\big\}\Big)\subseteq\operatorname*{argmin}_{y\in\mathcal{Y}(x,\xi)}\big\{\varphi(x,y,\xi)+\beta V(y)\big\}\neq\varnothing,$$

for all $(x,\xi)\in\mathcal{X}\times\Xi$.

Proof. (i): If $V_{\nu_\kappa}\xrightarrow{\text{e}}V$, then from Theorem 2.2 we deduce that the cluster point V must satisfy $V = B(V)$. With $\{V_{\nu_\kappa}\}_{\kappa\in\mathbb{N}}$ equi-lsc, it follows that $V_{\nu_\kappa}\xrightarrow{\text{p}}V$ as well [94, Theorem 7.10]. Since $\varphi_\nu\xrightarrow{\text{e}}\varphi$ and $\varphi_\nu\xrightarrow{\text{p}}\varphi$, for every $(x,\xi)\in\mathcal{X}\times\Xi$ it holds that $\varphi_\nu(x,\,\cdot\,,\xi)\xrightarrow{\text{e}}\varphi(x,\,\cdot\,,\xi)$ and $\varphi_\nu(x,\,\cdot\,,\xi)\xrightarrow{\text{p}}\varphi(x,\,\cdot\,,\xi)$. Reasoning similar to that in the proof of Lemma 2.1 as well as [96, Theorem 5.5(b)] then shows that the claimed inclusion holds. To see that the argmin is nonempty, observe that $\varphi(x,\,\cdot\,,\xi)+\beta V(\,\cdot\,)$ is lsc and real valued, and that $\mathcal{Y}(x,\xi)$ is nonempty and compact (as assumed throughout).

(ii): If $V = B(V)$ is unique, then $V_{\nu_\kappa}\xrightarrow{\text{e}}V$ for each epi-convergent subsequence of $\{V_\nu\}_{\nu\in\mathbb{N}}$. Viewing $\mathcal{V}\subseteq\text{lsc-fcns}(\mathcal{X})$ as a compact metric space under $d\!\!l$, it follows that $V_\nu\xrightarrow{\text{e}}V$. Similar reasoning as above then shows that the claimed inclusion and nonemptiness holds. □

2.4 Examples

In this section we illustrate the results previously developed in this chapter by applying them to the selling component of our supply chain problem from Example 1.1. While we only consider the consistent approximation of infinite-horizon stochastic optimal control problems via Theorem 2.2, similar arguments hold for finite-horizon problems and Theorem 2.1. In Subsection 2.4.1 we approximate a one-dimensional infinite-horizon problem and show that pointwise bounds on the solution functions to the approximating fixed-point problems (2.11) can be derived from problem primitives. This makes it possible to specify a set $\mathcal{V}\subseteq\text{lsc-fcns}(\mathcal{X})$ containing the approximating solutions for which we check that Theorem 2.2 applies. Next, in Subsection 2.4.2, we consider an alternate version of the previous problem for which Theorem 2.2 does not apply. We show that the resulting approximating decision rules fail to converge to an optimal policy for the true control problem. Finally, in Subsection 2.4.3 we approximate an infinite-horizon problem involving an autoregressive random variable, for which we adapt the approach taken in Subsection 2.4.1.

Throughout this section we work with saddle functions. For two convex sets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, a *saddle function* on $\mathcal{Z}_1\times\mathcal{Z}_2$ is a real-valued function which is convex in its first argument and

concave in its second argument, i.e., $f : \mathcal{Z}_1 \times \mathcal{Z}_2 \to \mathbb{R}$ such that $z_1 \mapsto f(z_1, z_2)$ is convex for each $z_2 \in \mathcal{Z}_2$ and $z_2 \mapsto f(z_1, z_2)$ is concave for each $z_1 \in \mathcal{Z}_1$.

2.4.1 Revenue Optimization with Stochastic Prices

Consider the infinite-horizon discrete-time stochastic optimal control problem

$$\begin{aligned} \underset{y^1, y^2, \ldots}{\text{minimize}} \quad & \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right], \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \boldsymbol{p}^t) \in [0, x^t], \quad t \in \mathbb{N}, \end{aligned} \tag{ROSP}$$

starting from an initial state $x^1 \in \mathbb{R}_+$. This is a one-dimensional version of our supply chain problem from Example 1.1 without the periodic-production aspect of the problem. Here $x^t \in \mathbb{R}_+$ is an inventory of some product, $\boldsymbol{p}^t \sim \mathbb{P} \in \mathfrak{P}(\mathbb{R}_+)$ is a random variable representing the per-unit price of the product on a spot market, and $C : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous convex function giving the cost of storing inventory between stages. We have to decide what inventory level to sell down to at the current realisation of the price.

The associated fixed-point problem is to find a function $V : \mathbb{R}_+ \to \mathbb{R}$ such that

$$V(x) = \mathbb{E}_{\mathbb{P}}\left[\inf_{0 \le y \le x} \Big\{C(y) - \boldsymbol{p}(x - y) + \beta V(y)\Big\}\right], \quad \text{for all } x \in \mathbb{R}_+.$$

With $\mathbb{P}$ supported on all of $\mathbb{R}_+$, the stage-cost function is unbounded. To approximate this problem, let $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots \sim \mathbb{P}$ be independent and identically distributed random variables, and suppose that $\{p_1, p_2, \ldots\}$ is an outcome of this random sequence. We set $\mathbb{P}_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} \mathbb{1}_{p_i}$ for each $\nu \in \mathbb{N}$ (as in SAA). Recall that $\mathbb{P}_\nu \Rightarrow \mathbb{P}$ for almost every such outcome $\{p_1, p_2, \ldots\}$ [20, Problem 3.1]. We do not approximate the stage-cost function. The approximating fixed-point problem is to find a function $V_\nu : \mathbb{R}_+ \to \mathbb{R}$ such that

$$V_\nu(x) = \mathbb{E}_{\mathbb{P}_\nu}\left[\inf_{0 \le y \le x} \Big\{C(y) - \boldsymbol{p}(x - y) + \beta V_\nu(y)\Big\}\right], \quad \text{for all } x \in \mathbb{R}_+.$$

Since $\mathbb{P}_\nu$ has compact support and future states are constrained to a compact set, an argument utilising [113, Theorems 9.6 and 9.8] shows that a unique continuous and convex solution V_ν satisfying the principle of optimality relative to the underlying approximation of (ROSP) exists.

In order to construct a set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathbb{R}_+)$ for which the conditions of Theorem 2.2 hold, we derive the following pointwise bound on the approximating fixed-point functions V_ν.

Proposition 2.3. *In the context of this subsection, suppose that $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ is finite. Then, for each approximation index $\nu \in \mathbb{N}$, we have that*

$$\inf_{\kappa\in\mathbb{N}}\Big\{-(1-\beta)^{-1}\cdot\mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}]x\Big\} \le V_\nu(x) \le (1-\beta)^{-1}\cdot C(x), \quad \textit{for all } x\in\mathbb{R}_+,$$

and these bounds are finite valued for almost every outcome $\{p_1, p_2, \ldots\}$.

Proof. For a given $\nu \in \mathbb{N}$, consider the underlying approximation of (ROSP) induced by $\mathbb{P}_\nu$:

$$\begin{aligned} \underset{y^1,y^2,\ldots}{\text{minimize}} \quad & \mathbb{E}_{\mathbb{P}_\nu^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}\Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right], \\ \text{subject to} \quad & x^{t+1} = y^t(x^t, \boldsymbol{p}^t) \in [0, x^t], \quad t\in\mathbb{N}. \end{aligned} \tag{2.14}$$

For each admissible control policy $y^1, y^2, \ldots$, we have

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}_\nu^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}\Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right] \quad \text{where } x^{t+1} = y^t(x^t,\boldsymbol{p}^t),\ t\in\mathbb{N} \\ \ge\ & \mathbb{E}_{\mathbb{P}_\nu^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}(-\boldsymbol{p}^t x^1)\right] \\ =\ & -(1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_\nu}[\boldsymbol{p}]x^1, \end{aligned}$$

the equality following from Lebesgue's dominated convergence theorem which applies since $\mathbb{P}_\nu$ is supported at a finite number of discrete points. Similarly,

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}_\nu^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}\Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right] \quad \text{where } x^{t+1} = y^t(x^t,\boldsymbol{p}^t),\ t\in\mathbb{N} \\ \le\ & \mathbb{E}_{\mathbb{P}_\nu^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}C(x^1)\right] \\ =\ & (1-\beta)^{-1}C(x^1). \end{aligned}$$

Since V_ν satisfies the principle of optimality relative to (2.14), we have that

$$\inf_{\kappa\in\mathbb{N}}\Big\{-(1-\beta)^{-1}\mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}]x\Big\} \le V_\nu(x) \le (1-\beta)^{-1}C(x), \quad \text{for all } x\in\mathbb{R}_+.$$

With $\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}]$ finite, the strong law of large numbers applies to $\mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}]$, and thus $\mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}] \to \mathbb{E}_{\mathbb{P}}[\boldsymbol{p}]$ for almost every outcome $\{p_1, p_2, \ldots\}$. Recalling that the infimum of a sequence of real numbers is finite if the sequence is convergent, it follows that for each $x \in \mathbb{R}_+$ the value

$$\inf_{\kappa \in \mathbb{N}} -(1-\beta)^{-1} \mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}] x$$

is finite for almost every outcome $\{p_1, p_2, \ldots\}$. □

In order to apply Theorem 2.2 we now construct a set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathbb{R}_+)$ that contains the approximating fixed-point functions. We assume that $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ is finite and use the pointwise bounds provided by Proposition 2.3. Let

$$\mathcal{V} \coloneqq \Big\{ V \in \text{lsc-fcns}(\mathbb{R}_+) : V \text{ is a convex function} \Big\}$$
$$\cap \Big\{ V \in \text{lsc-fcns}(\mathbb{R}_+) : \inf_{\kappa \in \mathbb{N}} \Big\{ -(1-\beta)^{-1} \cdot \mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}] x \Big\} \leq V(x) \leq (1-\beta)^{-1} \cdot C(x), \quad \text{for all } x \in \mathbb{R}_+ \Big\}$$

which contains the approximating fixed-point functions. As an intersection of closed sets, $\mathcal{V}$ is closed under dl [97, Propositions 4.1(i) and 4.5]. Furthermore, since each $V \in \mathcal{V}$ satisfies $V(0) = 0$, the point $z_{\text{ctr}} = (0, 0)$ is common to the epigraphs of every function in $\mathcal{V}$. It follows that $\mathcal{V}$ is bounded [95, Proposition 3.1] and therefore compact under dl [94, Proposition 4.45, Theorem 7.58].

We now check that Theorem 2.2 applies to $\mathcal{V}$. Consider Assumption 2.2 and the equi-lsc of epi-convergent sequences: for a given $V \in \mathcal{V}$ and $(x, p) \in \mathbb{R}_+ \times \mathbb{R}_+$, the value

$$b(V)(x,p) = \inf_{0 \leq y \leq x} \Big\{ C(y) - p(x-y) + \beta V(y) \Big\}$$

is finite, satisfying

$$\inf_{\kappa \in \mathbb{N}} \Big\{ -\Big(p + \beta(1-\beta)^{-1} \cdot \mathbb{E}_{\mathbb{P}_\kappa}[\boldsymbol{p}] \Big) x \Big\} \leq b(V)(x,p) \leq (1-\beta)^{-1} \cdot C(x). \tag{2.15}$$

Moreover, the mapping $p \mapsto b(V)(x,p)$ is concave. Hence, $b(V)(x, \cdot)$ is continuous and therefore measurable. Since the left- and right-hand side functions in (2.15) are integrable, we deduce that $b(V)(x, \cdot)$ is integrable as well. This shows that the operators B, B_ν, $\nu \in \mathbb{N}$ are well defined, and from (2.15) it can also be seen that they map $\mathcal{V}$ into itself. The equi-lsc of each epi-convergent sequence in $\mathcal{V}$ follows from convexity [97, Proposition 3.2(ii)] and (2.15).

Now consider the conditions (i) and (ii) of Theorem 2.2: The asymptotic semi-integrability statements (2.12) and (2.13) follow from the left- and right-hand functions in (2.15) being continuous and integrable. The equi-semicontinuity statements follow from $b(V)$ being a saddle function and (2.15); see [94, Example 9.14] and [97, Proposition 3.2(v)].

We thus conclude that Theorem 2.2 and Proposition 2.2(i) apply for almost every outcome $\{p_1, p_2, \ldots\}$. Also, if $V \in \mathcal{V}$ is a fixed point of B, then the principle of optimality can be verified using the pointwise bounds on each of the functions in $\mathcal{V}$; see [113, Theorem 9.2].

2.4.2 Failure to Converge to an Optimal Policy

We now study an example where the approximating decision rules do not converge to an optimal policy for the true control problem. Consider problem (ROSP) from Subsection 2.4.1, but with $\mathbb{P}$ being the Lévy$(0, 1)$ distribution having location parameter 0 and scale parameter 1; this is supported on all of $\mathbb{R}_+$ and has a heavy right-hand tail. In fact, $\mathbb{P}$ is not integrable and the reasoning of Subsection 2.4.1 does not apply.

Suppose that $C(0) = 0$, and take the policy that sells all inventory immediately regardless of the price, so that $y^1(x, p) = 0$ for all $(x, p) \in \mathbb{R}_+ \times \mathbb{R}_+$. When $x^1 > 0$, such a policy is trivially optimal for (ROSP), as

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right] \quad \text{where } x^{t+1} = 0,\ t \in \mathbb{N} \\ &= \mathbb{E}_{\mathbb{P}}[-\boldsymbol{p}^1 x^1] \\ &= -\infty. \end{aligned}$$

Following the scheme used in Subsection 2.4.1, we solve the approximating fixed-point problems numerically using `SDDP.jl` [38, 37] for random outcomes of $\{p_1, p_2, \ldots\}$. Setting $\beta = 0.99$ and $C(x) = \frac{1}{2}x^2$, Figure 2.1 graphs the decision $y_\nu(1, 1)$ as a function of ν, in which $y_\nu(1, 1) \to 1$ as $\nu \to \infty$.

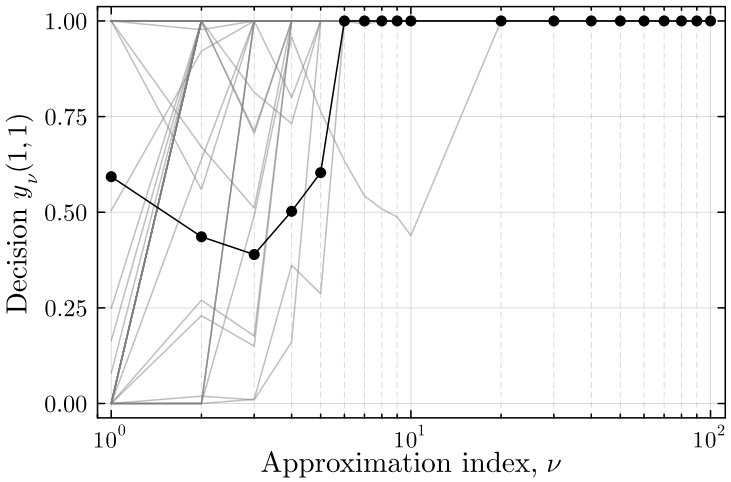

Figure 2.1: **Asymptotic Inconsistency.** Decision for the next state $y_\nu(1,1)$ as a function of the number of samples ν when the true distribution is Lévy$(0,1)$. One outcome highlighted.

It can further be shown that $y_\nu(x,p) \to x$ for each $(x,p) \in \mathbb{R}_+ \times \mathbb{R}_+$. The resulting policy is clearly suboptimal. Implementing it in (ROSP) yields the objective value

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \Big(C(x^{t+1}) - \boldsymbol{p}^t(x^t - x^{t+1})\Big)\right] \quad \text{where } x^{t+1} = x^t,\ t \in \mathbb{N} \\
&= (1-\beta)^{-1} \cdot C(x^1) \\
&> -\infty.
\end{aligned}$$

The failure of the approximating policy to converge to an optimal policy does not relate to a failure of the strong law of large numbers because $\mathbb{E}_{\mathbb{P}_\nu}[\boldsymbol{p}] \to \mathbb{E}_{\mathbb{P}}[\boldsymbol{p}] = \infty$ almost surely; see [21, Theorem 22.1] and its corollary. The first issue arises when attempting to construct a subset to contain the approximating fixed-point functions; the function previously used in Subsection 2.4.1 to specify a pointwise lower bound is not finite which does not enforce the required equi-semicontinuity properties. The second issue arises when attempting to verify the principle of optimality; even though it is satisfied for each $\nu \in \mathbb{N}$ relative to an underlying approximation of the true control problem, it does not apply asymptotically as $\nu \to \infty$.

2.4.3 Stagewise-Dependent Random Variables

Here we extend the control problem (ROSP) in Subsection 2.4.1 to model timewise dependence of the random spot price by expanding the state space. We suppose that the natural logarithm of the random price follows an autoregressive process: for current log price $\ell \in \mathbb{R}$ the log price in the next time period is the random variable $\alpha\ell + \boldsymbol{\xi}$, where $\alpha \in (0,1)$ and $\boldsymbol{\xi} \sim \mathbb{P} \in \mathfrak{P}(\mathbb{R})$. In the

next time period the previous log price ℓ becomes an additional state variable while the random variable $\boldsymbol{\xi}$ remains timewise independent, as in the formulation (SOC). The resulting fixed-point problem is to find a function $V : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ such that

$$V(x,\ell) = \mathbb{E}_{\mathbb{P}}\left[\inf_{\substack{0\le y\le x\\ \eta=\alpha\ell+\boldsymbol{\xi}}}\Big\{C(y) - \exp(\eta)\cdot(x-y) + \beta V(y,\eta)\Big\}\right], \quad \text{for all } (x,\ell)\in\mathbb{R}_+\times\mathbb{R}.$$

The possible realisations of $\alpha\ell + \boldsymbol{\xi}$ cannot be said to lie only in some compact subset of $\mathbb{R}$ when the distribution $\mathbb{P}$ has support on all of $\mathbb{R}$, so we now have an unbounded state variable and hence stage-cost function.

We do not assume that the constant α is available a priori and estimate it simultaneously to our approximation. Given a sequence of constants $\{\alpha_\nu \in (0,1)\}_{\nu\in\mathbb{N}}$ with $\alpha_\nu \to \alpha$ and a sequence of probability distributions $\{\mathbb{P}_\nu\}_{\nu\in\mathbb{N}}$ with $\mathbb{P}_\nu \Rightarrow \mathbb{P}$, the approximating fixed-point problem is to find a function $V_\nu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ such that

$$V_\nu(x,\ell) = \mathbb{E}_{\mathbb{P}_\nu}\left[\inf_{\substack{0\le y\le x\\ \eta=\alpha_\nu\ell+\boldsymbol{\xi}}}\Big\{C(y) - \exp(\eta)\cdot(x-y) + \beta V_\nu(y,\eta)\Big\}\right], \quad \text{for all } (x,\ell)\in\mathbb{R}_+\times\mathbb{R}.$$

Assuming that $\mathbb{P}_\nu$ has compact support, future states are constrained to a compact set. (Consider the future log prices when only the largest- or smallest-possible values of $\boldsymbol{\xi} \sim \mathbb{P}_\nu$ is realised.) An argument utilising [113, Theorem 9.6] shows that a unique continuous solution V_ν satisfying the principle of optimality relative to the underlying approximation of (ROSP) exists. Similar reasoning to that in the proof of [113, Theorem 9.8] shows that this V_ν is a saddle function.

Note that in the definition of V_ν the constraint $\eta = \alpha_\nu\ell + \boldsymbol{\xi}$ varies with ν. This is not explicitly permitted by the results previously developed in this chapter. However, we can treat this example by viewing the constant α_ν as a random variable that has the point-mass distribution $\mathbb{1}_{\alpha_\nu}$. Since $\alpha_\nu \to \alpha$, it holds that $\mathbb{1}_{\alpha_\nu} \Rightarrow \mathbb{1}_\alpha$, and with $\mathbb{P}_\nu \Rightarrow \mathbb{P}$ as well, $\mathbb{1}_{\alpha_\nu}\times\mathbb{P}_\nu \Rightarrow \mathbb{1}_\alpha\times\mathbb{P}$ [20, Theorem 2.8(ii)]. So we are in the setting of Section 2.3. In what follows, under absolute-integrability and light-tailedness assumptions, we verify that the assumptions of Theorem 2.2 do indeed hold.

We start by discussing how α_ν and $\mathbb{P}_\nu$ might arise in practice. Let $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots \sim \mathbb{P}$ be independent and identically distributed random variables, and suppose that $\{\xi^1, \xi^2, \ldots\}$ is an outcome of this random sequence. We can view historical log prices $\ell^1, \ell^2, \ldots$ as being generated from the formula $\ell^{t+1} = \alpha\ell^t + \xi^t$ and use this to construct estimates of the unknown parameter α.

For instance, applying ordinary least squares to the first $\nu+1$ log prices yields

$$\alpha_\nu \in \operatorname*{argmin}_{\bar\alpha,\bar\mu\in\mathbb{R}} \sum_{t=1}^{\nu} \big(\ell^{t+1} - (\bar\alpha\ell^t + \bar\mu)\big)^2.$$

The resulting α_ν can then be used to construct $\mathbb{P}_\nu$ by computing $\xi_\nu^t = \ell^{t+1} - \alpha_\nu \ell^t$ for each $t \in [\nu]$ and setting $\mathbb{P}_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} \mathbb{1}_{\xi_\nu^i}$. This differs from the standard SAA scheme in that the values of the already collected samples $\xi_\nu^1, \ldots, \xi_\nu^\nu$ vary with ν. Assuming that $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|]$ is finite, $\alpha_\nu \to \alpha$ for almost every outcome $\{\xi^1, \xi^2, \ldots\}$ [31, Section 3.3]. (Alternatives to least squares apply here as well.) Due to the following result, this also implies that $\mathbb{P}_\nu \Rightarrow \mathbb{P}$.

Proposition 2.4. *In the context of this subsection, suppose that $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|]$ is finite. If $\alpha_\nu \to \alpha$ for almost every outcome $\{\xi^1, \xi^2, \ldots\}$, then $\mathbb{P}_\nu \Rightarrow \mathbb{P}$ as well.*

Proof. Due to the strong law of large numbers,

$$\begin{aligned}
\frac{1}{\nu}\sum_{i=1}^{\nu} |\ell^i| &= \frac{1}{\nu}\sum_{i=1}^{\nu} \Big|\alpha^{i-1}\ell^1 + \alpha^{i-2}\xi^1 + \alpha^{i-1}\xi^2 + \cdots + \xi^{i-1}\Big| \\
&\le \frac{1}{\nu}\sum_{i=1}^{\nu} \Big(\alpha^{i-1}|\ell^1| + \alpha^{i-2}|\xi^1| + \alpha^{i-1}|\xi^2| + \cdots + |\xi^{i-1}|\Big) \\
&\to (1-\alpha)^{-1}\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|] \in \mathbb{R} \qquad (2.16)
\end{aligned}$$

for almost every outcome $\{\xi^1, \xi^2, \ldots\}$. Using the portemanteau theorem [75, Theorem 13.16(ii)], we show the following:

$$\mathbb{E}_{\mathbb{P}_\nu}[f(\boldsymbol{\xi})] = \frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi_\nu^i) \to \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})] \text{ for each bounded Lipschitz continuous } f : \Xi \to \mathbb{R},$$

for almost every outcome $\{\xi^1, \xi^2, \ldots\}$. Now, for each bounded L-Lipschitz continuous $f : \Xi \to \mathbb{R}$,

$$\left|\frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi^i) - \frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi_\nu^i)\right| \le \frac{1}{\nu}\sum_{i=1}^{\nu} \Big|f(\xi^i) - f(\xi_\nu^i)\Big| \le \frac{1}{\nu}\sum_{i=1}^{\nu} L|\xi^i - \xi_\nu^i| = \frac{1}{\nu}\sum_{i=1}^{\nu} L|(\alpha-\alpha_\nu)\ell^i|. \quad (2.17)$$

Since $\alpha_\nu \to \alpha$, using (2.16) we have that $\frac{1}{\nu}\sum_{i=1}^{\nu} L|(\alpha - \alpha_\nu)\ell^i| \to 0$, and we deduce from (2.17) that $\frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi^i)$ and $\frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi_\nu^i)$ have the same limit. With f bounded and Lipschitz continuous, $\frac{1}{\nu}\sum_{i=1}^{\nu} f(\xi^i) \to \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$ for almost every outcome $\{\xi^1, \xi^2, \ldots\}$. The result then follows from the transitivity of limits. □

As for Proposition 2.3 in Subsection 2.4.1, to construct a set $\mathcal{V} \subseteq \text{lsc-fcns}(\mathbb{R}_+ \times \mathbb{R})$ for which the conditions of Theorem 2.2 hold, we derive the following pointwise bound on the approximating fixed-point functions V_ν.

Proposition 2.5. *In the context of this subsection, suppose that* $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|]$ *and* $\mathbb{E}_{\mathbb{P}}[\exp(\boldsymbol{\xi})]$ *are finite. Then, for each* $\nu \in \mathbb{N}$ *we have that*

$$\inf_{\kappa\in\mathbb{N}}\left\{\sum_{t=1}^{\infty}\beta^{t-1}\cdot\Big(-\exp(\alpha_\kappa^t\ell)\cdot\textstyle\prod_{s=0}^{t-1}\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\alpha_\kappa^s\boldsymbol{\xi})]x\Big)\right\} \le V_\nu(x,\ell) \le (1-\beta)^{-1}\cdot C(x)$$

for all $(x,\ell)\in\mathbb{R}_+\times\mathbb{R}$,

and these bounds are finite valued for almost every outcome $\{\xi^1,\xi^2,\ldots\}$.

Proof. For a given $\nu \in \mathbb{N}$ and $(x,\ell) \in \mathbb{R}_+ \times \mathbb{R}$, similar reasoning to that in the proof of Proposition 2.3 shows

$$\begin{aligned}
V_\nu(x,\ell) &\ge \inf_{\kappa\in\mathbb{N}}\mathbb{E}_{\mathbb{P}_\kappa^\infty}\left[\sum_{t=1}^{\infty}\beta^{t-1}\Big(-\exp\big(\alpha_\kappa^t\ell+\alpha_\kappa^{t-1}\boldsymbol{\xi}^1+\alpha_\kappa^{t-2}\boldsymbol{\xi}^2+\cdots+\boldsymbol{\xi}^t\big)x\Big)\right]\\
&= \inf_{\kappa\in\mathbb{N}}\sum_{t=1}^{\infty}\beta^{t-1}\Big(-\exp(\alpha_\kappa^t\ell)\textstyle\prod_{s=0}^{t-1}\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\alpha_\kappa^s\boldsymbol{\xi})]x\Big) \qquad (2.18)\\
&\ge \inf_{\kappa\in\mathbb{N}}\sum_{t=1}^{\infty}\beta^{t-1}\Big(-(\alpha_\kappa^t\exp(\ell)+1)\textstyle\prod_{s=0}^{t-1}(\alpha_\kappa^s\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\boldsymbol{\xi})]+1)x\Big)\\
&\ge \sum_{t=1}^{\infty}\beta^{t-1}\Big(-\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^t\exp(\ell)+1)\textstyle\prod_{s=0}^{t-1}\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^s\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\boldsymbol{\xi})]+1)x\Big), \qquad (2.19)
\end{aligned}$$

where the equality follows from the properties of the exponential function and the monotone convergence theorem. With $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|]$ and $\mathbb{E}_{\mathbb{P}}[\exp(\boldsymbol{\xi})]$ finite, it can further be shown that (2.19) is an infinite sum of real numbers for almost every outcome $\{\xi^1,\xi^2,\ldots\}$. Applying the ratio test to successive terms yields

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{\beta^t\Big(-\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^{t+1}\exp(\ell)+1)\prod_{s=0}^{t}\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^s\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\boldsymbol{\xi})]+1)x\Big)}{\beta^{t-1}\Big(-\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^t\exp(\ell)+1)\prod_{s=0}^{t-1}\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^s\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\boldsymbol{\xi})]+1)x\Big)}\\
&= \limsup_{t\to\infty}\beta\sup_{\kappa\in\mathbb{N}}(\alpha_\kappa^t\mathbb{E}_{\mathbb{P}_\kappa}[\exp(\boldsymbol{\xi})]+1)\\
&= \beta\in(0,1),
\end{aligned}$$

which shows that the sum is convergent and hence that (2.18) is finite valued. □

Hence, assuming that the underlying $\mathbb{E}_{\mathbb{P}}[|\boldsymbol{\xi}|]$ and $\mathbb{E}_{\mathbb{P}}[\exp(\boldsymbol{\xi})]$ are finite, we set

$$\mathcal{V}_1 \coloneqq \left\{ V \in \text{lsc-fcns}(\mathbb{R}_+ \times \mathbb{R}) \ : \begin{array}{ll} V(x,\ell) \geq \inf\limits_{\kappa \in \mathbb{N}} \sum\limits_{t=1}^{\infty} \beta^{t-1} \cdot \Big(-\exp(\alpha_\kappa^t \ell) \cdot \prod_{s=0}^{t-1} \mathbb{E}_{\mathbb{P}_\kappa}[\exp(\alpha_\kappa^s \boldsymbol{\xi})] x \Big), & \\ V(x,\ell) \leq (1-\beta)^{-1} \cdot C(x), & \text{for all } (x,\ell) \in \mathbb{R}_+ \times \mathbb{R} \end{array} \right\},$$

which contains the approximating fixed-point functions. Compared to Subsection 2.4.1, some further work is required to construct a subset of lsc-fcns($\mathbb{R}_+ \times \mathbb{R}$) satisfying the equi-semicontinuity requirements of Theorem 2.2. We first show that the class of lsc saddle functions which are locally Lipschitz continuous with common moduli is closed under dl.

Proposition 2.6. *Let $\mathcal{Z}_1 \subseteq \mathbb{R}^{r_1}$ and $\mathcal{Z}_2 \subseteq \mathbb{R}^{r_2}$ be closed convex sets. For locally Lipschitz–continuous saddle functions $f_\nu \in \text{lsc-fcns}(\mathcal{Z}_1 \times \mathcal{Z}_2)$, $\nu \in \mathbb{N}$, with common modulus $L(z_1, z_2) < \infty$ at each $(z_1, z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$, if $f_\nu \xrightarrow{\text{e}} f \in \text{lsc-fcns}(\mathcal{Z}_1 \times \mathcal{Z}_2)$, then f is a saddle function that is locally Lipschitz continuous with modulus $L(z_1, z_2)$ at each $(z_1, z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$.*

Proof. Since each function in the sequence $\{f_\nu\}_{\nu \in \mathbb{N}}$ is locally Lipschitz continuous with common moduli, the sequence $\{f_\nu\}_{\nu \in \mathbb{N}}$ is equi-lsc on $\mathcal{Z}_1 \times \mathcal{Z}_2$ [97, Proposition 3.2(v)] and $f_\nu \xrightarrow{\text{e}} f$ implies that $f_\nu \xrightarrow{\text{p}} f$ [94, Theorem 7.10]. Thus, for every $z_2 \in \mathcal{Z}_2$, $\lambda \in [0,1]$, and $z_1, z_1' \in \mathcal{Z}_1$, we have

$$\begin{aligned} \lambda f(z_1, z_2) + (1-\lambda) f(z_1', z_2) &= \lim_{\nu \to \infty} \lambda f_\nu(z_1, z_2) + \lim_{\nu \to \infty} (1-\lambda) f_\nu(z_1', z_2) \\ &\geq \lim_{\nu \to \infty} f_\nu(\lambda z_1 + (1-\lambda) z_1', z_2) \\ &= f(\lambda z_1 + (1-\lambda) z_1', z_2), \end{aligned}$$

where the inequality follows from the fact that $z_1 \mapsto f_\nu(z_1, z_2)$ is convex. A similar argument shows that for every $z_1 \in \mathcal{Z}_1$, $\lambda \in [0,1]$, and $z_2, z_2' \in \mathcal{Z}_2$ the reverse inequality holds. Hence, f is a saddle function on $\mathcal{Z}_1 \times \mathcal{Z}_2$, and [97, Proposition 4.4] shows at each $(z_1, z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$ the function f is locally Lipschitz continuous with modulus $L(z_1, z_2)$. □

For almost every outcome $\{\xi^1, \xi^2, \ldots\}$, saddle functions in $\mathcal{V}_1$ are bounded on any open ball in the interior of their domain via the continuous pointwise bounds on the functions in $\mathcal{V}_1$. It follows that each saddle function in $\mathcal{V}_1$ is locally Lipschitz continuous with common modulus function $L : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ [94, Example 9.14]. Let

$$\mathcal{V}_2 \coloneqq \Big\{ V \in \text{lsc-fcns}(\mathbb{R}_+ \times \mathbb{R}) : V \text{ is an } L\text{-locally Lipschitz–continuous saddle function} \Big\},$$

which contains each of the approximating fixed-point functions. Proposition 2.6 shows that $\mathcal{V}_2$ is closed under $d\!\!l$.

We set $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2$ in order to apply Theorem 2.2. The set $\mathcal{V}$ is closed under $d\!\!l$, and since each $V \in \mathcal{V}$ satisfies $V(0,0) = 0$, the point $z_{\text{ctr}} = (0,0,0)$ shows that $\mathcal{V}$ is bounded and therefore compact under $d\!\!l$. Assumption 2.2, the equi-lsc of epi-convergent sequences, and the conditions (i) and (ii) of Theorem 2.2, can all be verified using similar reasoning as in Subsection 2.4.1. We thus conclude that Theorem 2.2 and Proposition 2.2(i) apply for almost every outcome $\{\xi^1, \xi^2, \ldots\}$. Also, if $V \in \mathcal{V}$ is a fixed point of B, then the principle of optimality can be verified using the pointwise bounds on the functions in $\mathcal{V}$; see [113, Theorem 9.2].

2.5 Discussion

In this chapter we study the convergence of approximations of stochastic dynamic programs. Up to a high level of generality, consistency depends upon the existence of a set of functions under which the approximating Bellman operators satisfy certain asymptotic integrability and semicontinuity conditions. This generality allows us to treat schemes where cost functions (and even constraints after suitable reformulations) are simultaneously approximated. Although the conditions are somewhat abstract, in two different examples we demonstrate that they can be verified relatively easily by exploiting problem-specific convexity and concavity properties.

The analysis in this chapter is restricted to *in*-sample convergence; for a sequence of approximating distributions, we consider only the value functions they induce and the limiting forms of these value functions. In practice however, decision-making policies derived from approximating value functions are followed *out of* sample. Although reassuring that an approximation scheme asymptotically converges, this is not an out-of-sample performance guarantee for a specific decision-making policy obtained at a particular level of approximation.

Chapter 3

Stochastic Dynamic Programming and Model Predictive Control

In the previous chapter we showed that sample-based approximations of stochastic optimal control problems are asymptotically consistent under fairly general conditions. This is conceptual support for the belief that the decisions obtained from solving these approximating problems will perform well in practice, as long as the sample size is large enough. Although one can use 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 [110, 104] and solving the resulting problem is computationally expensive [44].[1] It follows that, for large problems, SAA may only be practically applied when the number of samples in each stage is small, risking overfitting to the in-sample data and poor performance *out of sample*.

In this chapter we turn to studying performance out of sample, being interested in different data-driven approximations in the small-sample regime. In principle, such problems are amenable to solution via stochastic dynamic programming (SDP), as long as the state dimension is not too high [13]. However, often model predictive control (MPC) is used as an alternative. Recall that MPC fixes the random variables involved in the stochastic optimal control problem to deterministic forecasts, giving a cost-to-go function that is computationally inexpensive to evaluate [14]. In this way MPC introduces a further aspect of structural approximation to the problem, and this enables it to handle high state dimensions, nonlinear constraints, and a large number of stages.

[1] In certain circumstances there may however be alternative distributional approximation schemes that are less affected by the number of stages.

We take a sample-based approach to both SDP and MPC, using SAA to construct distributional approximations via N random samples from the true probability distribution. In the case of MPC, we consider a simple formulation that forecasts the random variables as the mean of these N samples. Compared to SDP, in certain applications MPC performs poorly out of sample [86], and in other applications MPC performs well out of sample [84]. This is surprising as MPC is tantamount to ignoring randomness. The intention of this chapter is to advance our understanding of what problem characteristics drive this performance difference, by comparing out-of-sample performance when N is small.

This chapter is based on the joint work:

On the Out-of-Sample Performance of
Stochastic Dynamic Programming and Model Predictive Control
Dominic S. T. Keehan, Andrew B. Philpott, & Edward J. Anderson [70].

In this chapter we consider only SDP and MPC, but there are many alternative structural-approximation schemes that could also perform well. Emblematic of a wider family of sample-based decomposition algorithms, the celebrated stochastic *dual* dynamic programming algorithm, which iteratively refines piecewise-linear approximations of solutions to the sample-based SDP functional equations, can be very efficient when $\mathcal{Y}$ is convex valued and φ is a convex function [90]. However, this is not so much the case when φ is a saddle function (as for some formulations of our supply chain problem from Example 1.1) [35], so these algorithms are still limited to moderate sample sizes in general. Under certain conditions, such sample-based decomposition algorithms converge to SDP as the piecewise-linear approximations are refined [91, 54]. Thus, for an implementation with sufficient run time, the out-of-sample performance of the resulting decision policy would be similar to, and hence can be studied through, that of SDP.

Another alternative approximation scheme could use a reinforcement-learning approach, where state and action pairs are observed and a suitable decision policy learned over time. But this would significantly complicate our analysis, so we do not consider such an approach in this thesis. Do note however that MPC shares some conceptual similarities with reinforcement learning through the updating of forecasts as state and action pairs are observed [15], and this is akin to the forecasting approach that we develop in Chapter 6 for the New Zealand dairy supply chain problem.

To understand the out-of-sample performance differences of SDP and MPC, we first look through the lens of distributional ambiguity. In data-driven stochastic optimization, when the number of samples N is small, considering an ambiguity set of probability distributions informed by the N samples can improve out-of-sample performance [6, 57]. Moment constraints are a classical choice for constructing these ambiguity sets [100, 42, 33]. 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* rigorous 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, and therefore when MPC is a better choice of approximation in practice, we study a specific stochastic revenue optimization problem from the supply chain class of Example 1.1 that determines how much product from an existing inventory to sell at a given price offered by a spot 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 stochastic 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.

The remainder of this chapter is organised as follows. In Section 3.1 we formulate a general stochastic optimal control problem as well as its data-driven approximations via sample-based SDP and MPC. In Section 3.2 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 3.3 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 3.4 we introduce our stochastic 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 rest of Section 3.4 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 chapter concludes with a discussion in Section 3.5.

3.1 Stochastic Optimal Control

Recall that, given a closed state space $\mathcal{X} \subseteq \mathbb{R}^n$ and outcome space $\Xi \subseteq \mathbb{R}^m$, we consider infinite-horizon stochastic optimal control problems of the form

$$\begin{array}{ll} \underset{y_1, y_2, \ldots}{\text{minimize}} & \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \\ \text{subject to} & x_{t+1} = y_t(x_t, \boldsymbol{\xi}_t) \in \mathcal{Y}(x_t, \boldsymbol{\xi}_t), \quad t \in \mathbb{N}, \end{array} \tag{SOC}$$

starting from an initial state $x_1 \in \mathcal{X}$. The stagewise-independent random vector $\boldsymbol{\xi}_t \sim \mathbb{P} \in \mathfrak{P}(\Xi)$ for each $t \in \mathbb{N}$, and the infinite-product distribution $\mathbb{P}^\infty := \mathbb{P} \times \mathbb{P} \times \cdots$ defines a joint distribution for the random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$. Here the discount factor $\beta \in (0, 1)$, and the stage-cost function $\varphi : \mathcal{X} \times \mathcal{X} \times \Xi \to \mathbb{R}$. The set-valued mapping $\mathcal{Y} : \mathcal{X} \times \Xi \rightrightarrows \mathcal{X}$ defines admissible policies $\{y_1, y_2, \ldots : \mathcal{X} \times \Xi \to \mathcal{X}\}$, and we further require the mappings $\xi \mapsto y_t(x, \xi)$, $t \in \mathbb{N}$, to be measurable for every $x \in \mathcal{X}$. In this chapter 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 [18, Chapters 3 and 4].

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

Assumption 3.1.

(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 $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ 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 3.1(ii) does not preclude the modelling of stagewise-dependent randomness, since auxiliary states that track the realisations of random vectors in previous stages can be created, as we do in Example 3.1 below. Moreover, Assumption 3.1(iii) can often be relaxed, although the mathematical tools required become more complex.

Recall that a problem closely related to (SOC) is that of finding a real-valued solution function $v : \mathcal{X} \times \Xi \to \mathbb{R}$ to the SDP functional equation

$$v(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{\xi})\big] \Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi. \tag{3.1}$$

Under Assumption 3.1, the equation (3.1) has a unique bounded and continuous solution v. Moreover, the value $\mathbb{E}_{\mathbb{P}}[v(x_1, \boldsymbol{\xi})]$ 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

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

This defines a decision rule at each stage and the resulting policy solves (SOC). For a discussion of all of these well-known results, see, e.g., [113, Chapter 9], or [18, 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).

3.1.1 Data-Driven Stochastic Optimal Control

To approximate (SOC) in a data-driven manner, given N samples $\xi_1, \ldots, \xi_N \in \Xi$ believed to be drawn from $\mathbb{P}$, we form the empirical distribution $\mathbb{P}_N \coloneqq \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\xi_i} \in \mathfrak{P}(\Xi)$. 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

$$\begin{aligned} \underset{y_1, y_2, \ldots}{\text{minimize}} \quad & \mathbb{E}_{\mathbb{P}_N^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \\ \text{subject to} \quad & x_{t+1} = y_t(x_t, \boldsymbol{\xi}_t) \in \mathcal{Y}(x_t, \boldsymbol{\xi}_t), \quad t \in \mathbb{N}. \end{aligned} \tag{3.2}$$

Just as for (3.1), this is closely related to the problem of finding a real-valued solution function $v_{\mathrm{S}} : \mathcal{X} \times \Xi \to \mathbb{R}$ to the sample-based SDP functional equation

$$v_{\mathrm{S}}(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta \mathbb{E}_{\mathbb{P}_N}\big[v_{\mathrm{S}}(y, \boldsymbol{\xi})\big] \Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi. \tag{3.3}$$

Under Assumption 3.1 the same remarks made with reference to (3.1) apply to (3.3), and by solving (3.3) a sample-based SDP policy $y_{\mathrm{S}} : \mathcal{X} \times \Xi \to \mathcal{X}$ satisfying

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

can be obtained and applied out of sample to the true problem (SOC).

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

$$v_{\mathrm{M}}(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta v_{\mathrm{M}}(y, \mu_N) \Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi. \tag{3.4}$$

Observe that (3.4) is equivalent to (3.1) with the point-mass distribution $\mathbb{1}_{\mu_N} \in \mathfrak{P}(\Xi)$ used instead of $\mathbb{P}$. Under Assumption 3.1 the same remarks regarding uniqueness, boundedness, continuity, and measurability made with reference to (3.1) apply to (3.4), and by solving (3.4) a sample-based MPC policy $y_{\mathrm{M}} : \mathcal{X} \times \Xi \to \mathcal{X}$ satisfying

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

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).

3.2 Connections with Distributionally Ambiguous Optimization

Recall that the SDP policy obtained by solving (3.3) depends critically on the values of the N samples used. Instead, by hedging against the worst-case distributions in an *ambiguity set* $\mathcal{P} \subseteq \mathfrak{P}(\Xi)$ informed by the N samples, *distributionally robust optimization* (*or* DRO) provides a means to limit out-of-sample disappointment. The infinite product of these ambiguity sets $\mathcal{P}^\infty \coloneqq \mathcal{P} \times \mathcal{P} \times \cdots$ defines a set of joint distributions for the random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$, which yields the distributionally robust problem

$$\begin{array}{ll} \underset{y_1, y_2, \ldots}{\text{minimize}} & \sup\limits_{\mathbb{Q} \in \mathcal{P}^\infty} \mathbb{E}_{\mathbb{Q}} \left[\sum\limits_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t) \right] \\ \text{subject to} & x_{t+1} = y_t(x_t, \boldsymbol{\xi}_t) \in \mathcal{Y}(x_t, \boldsymbol{\xi}_t), \quad t \in \mathbb{N}. \end{array} \tag{3.5}$$

The choice of $\mathcal{P}^\infty$ in (3.5) requires care. By using a product set, under the assumption of stage-wise independence, $\mathcal{P}^\infty$ satisfies the "rectangularity" property (see [67, Assumption 2.1] and [105] for details), whereby (3.5) can be solved by finding a real-valued solution function $v_{\mathrm{R}} : \mathcal{X} \times \Xi \to \mathbb{R}$ to the DRO functional equation

$$v_{\mathrm{R}}(x, \xi) = \inf_{y \in \mathcal{Y}(x,\xi)} \Big\{ \varphi(x, y, \xi) + \beta \cdot \sup\nolimits_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}} \big[v_{\mathrm{R}}(y, \boldsymbol{\xi}) \big] \Big\}, \quad \text{for all } (x, \xi) \in \mathcal{X} \times \Xi. \tag{3.6}$$

Note that the ambiguity set in (3.6) 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 (3.5).

To demonstrate a connection between DRO and MPC, we now analyse the functional operator associated with (3.4). 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 3.1 the equation (3.4) 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

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

The operator B_{M} can be defined for any forecast in place of the sample average μ_N appearing here. For a proof of the fact that B_{M} maps $\mathcal{C}(\mathcal{X} \times \Xi)$ to itself, see, e.g., [113, Theorem 9.6].

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

$$\left\|B_{\mathrm{M}}(f) - B_{\mathrm{M}}(g)\right\|_{\infty} \leq \beta \|f - g\|_{\infty}, \quad \text{for all } f, g \in \mathcal{C}(\mathcal{X} \times \Xi),$$

and $\beta \in (0, 1)$. Banach's fixed-point theorem then shows that the equation $f = B_{\mathrm{M}}(f)$ has a unique solution in $(\mathcal{C}(\mathcal{X} \times \Xi), \|\cdot\|_{\infty})$, and this solution solves the MPC functional equation (3.4). Further, if there exists a closed subset $\mathcal{F} \subseteq \mathcal{C}(\mathcal{X} \times \Xi)$ which maps to itself under B_{M} and $f \in \mathcal{C}(\mathcal{X} \times \Xi)$ solves $f = B_{\mathrm{M}}(f)$, then we have that $f \in \mathcal{F}$ [113, Theorem 3.2, Corollary 1]. This motivates the following definition.

Definition 3.1. The MPC Bellman operator B_{M} is *concavity preserving* (respectively *convexity preserving*) if there exists a closed subset $\mathcal{F} \subseteq \mathcal{C}(\mathcal{X} \times \Xi)$ where the mapping $\xi \mapsto f(x, \xi)$ is concave (convex) for all $x \in \mathcal{X}$ 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 3.1. *Let Assumption 3.1 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 all $x, y \in \mathcal{X}$. Then the MPC Bellman operator B_{M} is concavity preserving.*

Proof. Consider the set

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

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

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

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

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 the mapping $\xi \mapsto B_{\mathrm{M}}(f)(x, \xi)$ is concave for all $x \in \mathcal{X}$, and hence that $B_{\mathrm{M}}(f) \in \mathcal{F}$. □

In some circumstances we can establish this key property of concavity preservation without requiring all of the conditions of Proposition 3.1. Some special cases will be discussed in the examples below.

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

Theorem 3.1. *Let Assumption 3.1 hold. For an ambiguity set $\mathcal{P} \subseteq \mathfrak{P}(\Xi)$, suppose that $\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mu_N$ for all $\mathbb{Q} \in \mathcal{P}$ and $\mathbb{1}_{\mu_N} \in \mathcal{P}$. If the MPC Bellman operator B_{M} is concavity preserving, then the DRO functional equation (3.6) has a solution which is the same as the solution to the MPC functional equation (3.4).*

Proof. The equation (3.6) 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

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

With B_{M} concavity preserving, there exists a closed subset $\mathcal{F} \subseteq \mathcal{C}(\mathcal{X} \times \Xi)$ where the mapping $\xi \mapsto f(x, \xi)$ is concave for all $x \in \mathcal{X}$ 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 each $f \in \mathcal{F}$ we have that

$$
\mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] \leq f(y, \mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}]) = f(y, \mu_N) \quad \text{for all } \mathbb{Q} \in \mathcal{P}, \tag{3.7}
$$

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 (3.7) implies $\sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\big[f(y, \boldsymbol{\xi})\big] \leq f(y, \mu_N)$.

But, with $\mathbb{E}_{\mathbb{1}_{\mu_N}}[f(y,\boldsymbol{\xi})] = f(y,\mu_N)$ and $\mathbb{1}_{\mu_N} \in \mathcal{P}$, this supremum is attained by $\mathbb{1}_{\mu_N}$ and $\sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[f(y,\boldsymbol{\xi})] = f(y,\mu_N)$. Thus, for all $f \in \mathcal{F}$, we have that

$$\begin{aligned} B_{\mathrm{R}}(f)(x,\xi) &= \inf_{y\in\mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \sup\nolimits_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[f(y,\boldsymbol{\xi})] \Big\} \\ &= \inf_{y\in\mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta f(y,\mu_N) \Big\} = B_{\mathrm{M}}(f)(x,\xi), \quad \text{for all } (x,\xi) \in \mathcal{X}\times\Xi, \end{aligned} \tag{3.8}$$

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 (3.6) has a solution which is the same as the solution to the MPC functional equation (3.4). 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 (3.6). In light of (3.8), this solution is the same as the solution to (3.4). □

In contrast to DRO, *distributionally optimistic optimization* (*or* DOO) considers only the best-case distributions in an ambiguity set $\mathcal{P} \subseteq \mathfrak{P}(\Xi)$. Using the rectangular ambiguity set $\mathcal{P}^\infty$, a distributionally optimistic version of (3.5) can be formed. This leads to the problem of finding a real-valued solution function $v_{\mathrm{O}} : \mathcal{X}\times\Xi \to \mathbb{R}$ to the DOO functional equation

$$v_{\mathrm{O}}(x,\xi) = \inf_{y\in\mathcal{Y}(x,\xi)} \Big\{ \varphi(x,y,\xi) + \beta \cdot \inf\nolimits_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}[v_{\mathrm{O}}(y,\boldsymbol{\xi})] \Big\}, \quad \text{for all } (x,\xi) \in \mathcal{X}\times\Xi. \tag{3.9}$$

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, then the following converse result to Theorem 3.1 holds.

Corollary 3.1. *Suppose that the conditions in the statement of Theorem 3.1 hold, but that the MPC Bellman operator B_{M} is convexity preserving rather than concavity preserving. Then the DOO functional equation (3.9) has a solution which is the same as the solution to the MPC functional equation (3.4).*

We now give some examples of problems with concavity- and convexity-preserving MPC Bellman operators that differ from those characterised by Proposition 3.1.

Example 3.1. Downward, Dowson, & Baucke [35] study a class of problems where stagewise-dependent uncertainty appears in the objective function (see also [36, Chapter 4]). These

include resource allocation problems under price uncertainty, such as hydroelectric bidding and agricultural production planning, as well as our supply chain problem from Example 1.1. 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

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

where

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

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

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

For each $f \in \mathcal{F}$ 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 3.1. Thus, sample-based MPC satisfies the DRO interpretation of Theorem 3.1. □

Example 3.2. The *hydrothermal scheduling problem* [90] 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 $\boldsymbol{\xi} \in \Xi$. All quantities are measured in terms of energy. The functional equation is

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

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 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 each $f \in \mathcal{F}$ 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 3.1 (although with the inequalities reversed). Thus, sample-based MPC satisfies the DOO interpretation of Corollary 3.1. □

If the MPC Bellman operator is concavity preserving, then Theorem 3.1 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 reduced decision variance [6]. Conversely, if the MPC Bellman operator is convexity preserving, then Corollary 3.1 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 overfitting and increased decision variance [57]. This suggests that, for small sample sizes, when the MPC Bellman operator is concavity preserving as in Example 3.1, MPC may be a better choice than SDP. On the other hand, when the MPC Bellman operator is convexity preserving as in Example 3.2, SDP may be a better choice than MPC. In the following section we provide performance guarantees which support these observations.

3.3 Out-of-Sample Performance

Suppose that the distribution $\mathbb{P}$ has a finite mean $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$. For N independent and identically distributed 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 that $\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\xi}_i \to \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ almost surely as $N \to \infty$. MPC can take advantage of this situation since the computational complexity of solving (3.4) 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 (3.4) 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 3.2. *Let Assumption 3.1 hold and set $\mu_N = \mu \coloneqq \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}]$ in the MPC functional equation (3.4). If the resulting MPC Bellman operator B_{M} is concavity preserving, then*

$$\mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right] \leq v_{\mathrm{M}}(x_1, \mu), \quad \text{where } x_{t+1} = y_{\mathrm{M}}(x_t, \boldsymbol{\xi}_t),\ t \in \mathbb{N}.$$

That is, the value $v_{\mathrm{M}}(x_1, \mu)$ obtained by solving the MPC functional equation (3.4) is an upper bound on the cost incurred when applying the resulting policy to (SOC).

Proof. Set $\mathcal{P} = \{\mathbb{P}, \mathbb{1}_\mu\}$, and recall that $\mathcal{P}^\infty = \mathcal{P} \times \mathcal{P} \times \cdots$ is a rectangular set. By Theorem 3.1, the MPC policy solves (3.5), and we have that

$$\begin{aligned} v_{\mathrm{M}}(x_1, \mu) &= \inf_{y_1, y_2, \cdots} \sup_{\mathbb{Q} \in \mathcal{P}^\infty} \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \quad \text{subject to } x_{t+1} = y_t(x_t, \boldsymbol{\xi}_t),\ t \in \mathbb{N} \\ &= \sup_{\mathbb{Q} \in \mathcal{P}^\infty} \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{\infty} \beta^{t-1} \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \quad \text{where } x_{t+1} = y_{\mathrm{M}}(x_t, \boldsymbol{\xi}_t),\ t \in \mathbb{N}. \end{aligned} \tag{3.10}$$

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

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

which holds since $\mathbb{P}^\infty \in \mathcal{P}^\infty$. The right-hand side equals $v_{\mathrm{M}}(x_1, \mu)$ by virtue of (3.10). □

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. In fact, viewing $\mathbb{1}_\mu$ as the underlying distribution being approximated, since $\mathbb{1}_{\mu_N} \Rightarrow \mathbb{1}_\mu$ almost surely, the epi-consistency results of the previous Chapter 2 can be applied here as well. 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 3.2. *Suppose that the conditions in the statement of Theorem 3.2 hold, but that the MPC Bellman operator is convexity preserving rather than concavity preserving. Then*

$$\mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right] \geq v_{\mathrm{M}}(x_1, \mu), \quad \text{where } x_{t+1} = y_{\mathrm{M}}(x_t, \boldsymbol{\xi}_t),\ t \in \mathbb{N}.$$

That is, the value $v_{\mathrm{M}}(x_1, \mu)$ obtained by solving the MPC functional equation (3.4) is a lower bound on the cost incurred when applying the resulting policy to (SOC).

Proof. The proof repeats the proof of Theorem 3.2, but with the reference to Theorem 3.1 replaced by a reference to Corollary 3.1, the supremum operations replaced by infimum operations, and the inequalities reversed. □

Corollary 3.2 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 problem (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. We denote this value as a function of the initial state by

$$\widetilde{V}(x_1; y) \coloneqq \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \varphi(x_t, x_{t+1}, \boldsymbol{\xi}_t)\right], \quad \text{where } x_{t+1} = y(x_t, \boldsymbol{\xi}_t),\ t \in \mathbb{N}. \tag{3.11}$$

If y is constructed from samples, then $\widetilde{V}(x_1; y)$ is the *out-of-sample cost* of y.

To study out-of-sample performance, we do not require the total-boundedness condition of Assumption 3.1(iii). Instead, we make the following integrability assumption.

Assumption 3.2. *Assumption 3.1(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 $g(\boldsymbol{\xi})$ with $\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{\xi})] < \infty$, such that for every $x, y \in \mathcal{X}$ and $\mathbb{P}$-almost every $\xi \in \Xi$, we have that $g(\xi) \geq |\varphi(x, y, \xi)|$.*

Under Assumption 3.2(iii), the t-th term in the sum in (3.11) is bounded by the random variable $g(\boldsymbol{\xi}_t)$ which has a finite expectation. With $\beta \in (0, 1)$, it follows that $\widetilde{V}$ is bounded. Having defined $\widetilde{V}$ as a function of the initial state and shown that it is bounded, it satisfies the functional equation

$$\widetilde{V}(x; y) = \mathbb{E}_{\mathbb{P}}\Big[\varphi(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}\big(y(x, \boldsymbol{\xi}); y\big)\Big], \quad \text{for all } x \in \mathcal{X}. \tag{3.12}$$

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 3.2. For two admissible policies $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, define

$$\widetilde{V}_T(x; y, y') \coloneqq \mathbb{E}_{\mathbb{P}}\Big[\varphi(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}_{T-1}(y(x, \boldsymbol{\xi}); y, y')\Big], \quad \text{for all } T \in \mathbb{N},$$

where $\widetilde{V}_0(x; y, y') \coloneqq \widetilde{V}(x; y')$.

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

Lemma 3.1. *Let Assumption 3.2 hold. Then for two admissible policies $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, we have that $\lim_{T\to\infty} |\widetilde{V}_T(x; y, y') - \widetilde{V}(x; y)| = 0$ for all $x \in \mathcal{X}$.*

Proof. For $x_1 \in \mathcal{X}$ observe that

$$\begin{aligned}
&\widetilde{V}_1(x_1; y, y') - \widetilde{V}(x_1; y) \\
&= \mathbb{E}_{\mathbb{P}}\Big[\varphi(x_1, y(x_1, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}(y(x_1, \boldsymbol{\xi}); y')\Big] - \mathbb{E}_{\mathbb{P}}\Big[\varphi(x_1, y(x_1, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}(y(x_1, \boldsymbol{\xi}); y)\Big] \\
&= \beta \mathbb{E}_{\mathbb{P}}\Big[\widetilde{V}(y(x_1, \boldsymbol{\xi}); y') - \widetilde{V}(y(x_1, \boldsymbol{\xi}); y)\Big].
\end{aligned}$$

Continuing this reasoning, an induction shows that

$$\Big|\widetilde{V}_T(x_1; y, y') - \widetilde{V}(x_1; y)\Big| = \beta^T \Big|\mathbb{E}_{\mathbb{P}^\infty}\left[\widetilde{V}(x_T; y') - \widetilde{V}(x_T; y)\right]\Big|, \quad \text{where } x_{t+1} = y(x_t, \boldsymbol{\xi}_t),\ t \in [T-1]. \tag{3.13}$$

Since $\widetilde{V}$ is bounded under Assumption 3.2 and $\beta^T \to 0$ as $T \to \infty$, the value of (3.13) vanishes as $T \to \infty$. Replacing x_1 with x concludes the proof. □

Theorem 3.3. *Let Assumption 3.2 hold. For two admissible policies $y, y' : \mathcal{X} \times \Xi \to \mathcal{X}$, if $\widetilde{V}(x; y') \le \widetilde{V}_1(x; y, y')$ for all $x \in \mathcal{X}$, then $\widetilde{V}(x; y') \le \widetilde{V}(x; y)$ for each $x \in \mathcal{X}$.*

Proof. Writing $\widetilde{V}(x; y')$ as $\widetilde{V}_0(x; y, y')$, we pose the inductive hypothesis $\widetilde{V}_{t-1}(x; y, y') \le \widetilde{V}_t(x; y, y')$ for all $x \in \mathcal{X}$, which is true for $t = 1$ by assumption. By the hypothesis, we have $\widetilde{V}_{t-1}(y(x, \xi); y, y') \le \widetilde{V}_t(y(x, \xi); y, y')$, and it follows that

$$\begin{aligned}
\widetilde{V}_t(x; y, y') &= \mathbb{E}_{\mathbb{P}}\Big[\varphi(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}_{t-1}(y(x, \boldsymbol{\xi}); y, y')\Big] \\
&\le \mathbb{E}_{\mathbb{P}}\Big[\varphi(x, y(x, \boldsymbol{\xi}), \boldsymbol{\xi}) + \beta \widetilde{V}_t(y(x, \boldsymbol{\xi}); y, y')\Big] = \widetilde{V}_{t+1}(x; y, y'),
\end{aligned} \tag{3.14}$$

for all $x \in \mathcal{X}$. The inequality (3.14) establishes the induction. Hence, $\widetilde{V}(x; y') \leq \widetilde{V}_t(x; y, y')$ for each $t \in \mathbb{N}$, and thus $\widetilde{V}(x; y') \leq \lim_{t\to\infty} \widetilde{V}_t(x; y, y') = \widetilde{V}(x; y)$ via Lemma 3.1. □

Rather than having to calculate integrals directly to compare the performance of different policies, Theorem 3.3 allows us to check a uniform condition involving similarly defined functional equations.

3.4 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 3.1. This problem is an example from the class for which sample-based MPC satisfies the DRO interpretation of Theorem 3.1. Recall the stochastic revenue optimization problem from Section 2.4 of the previous chapter, i.e., the infinite-horizon discrete-time stochastic optimal control problem

$$\begin{aligned} \underset{y_1, y_2, \cdots}{\text{maximize}} \quad & \mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\infty} \beta^{t-1} \cdot \Big(\boldsymbol{p}_t(x_t - x_{t+1}) - C(x_{t+1})\Big)\right], \\ \text{subject to} \quad & x_{t+1} = y_t(x_t, \boldsymbol{p}_t) \in [0, x_t], \quad t \in \mathbb{N}, \end{aligned} \tag{ROSP}$$

where we start from an initial state $x_1 \in \mathbb{R}_+$. This is a one-dimensional version of our supply chain problem from Example 1.1 without the periodic-production aspect of the problem. Here $x_t \in \mathbb{R}_+$ is an inventory of some product, $\boldsymbol{p}_t \sim \mathbb{P} \in \mathfrak{P}(\mathbb{R})$ is a random variable representing the per-unit price of the product on a spot market, and $C : \mathbb{R}_+ \to \mathbb{R}_+$ is a function giving the cost of storing inventory between time periods. The problem is to decide what inventory level to sell down to at the current realisation of the price. We assume that C is increasing, strictly convex, and continuously differentiable. Writing $c(x) \coloneqq \frac{d}{dx}C(x)$, 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 $\lim_{x\to\infty} c(x) = \infty$.

Beyond our supply chain context, problem (ROSP) is that facing a decision maker who maximizes their expected total 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.

Just as for (SOC), problem (ROSP) is closely related to that of finding a real-valued solution function $v : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ to the SDP functional equation

$$v(x,p) = \sup_{0 \le y \le x} \Big\{ p(x-y) - C(y) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{p})\big] \Big\}, \quad \text{for all } (x,p) \in \mathbb{R}_+ \times \mathbb{R}. \tag{3.15}$$

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 (3.15) has a unique continuous solution v and that the mapping $x \mapsto v(x,p)$ is concave for all $p \in \mathbb{R}$ [113, Chapter 9]. Given such v, a policy $y : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ satisfying

$$y(x,p) \in \operatorname*{argmax}_{0 \le y \le x} \Big\{ p(x-y) - C(y) + \beta \mathbb{E}_{\mathbb{P}}\big[v(y, \boldsymbol{p})\big] \Big\}, \quad \text{for all } (x,p) \in \mathbb{R}_+ \times \mathbb{R}, \tag{3.16}$$

is optimal for (ROSP).

In the notation of Section 3.1, (ROSP) has $\mathcal{Y}(x,p) = [0,x]$, and a suitable change of perspective from maximization to minimization yields the stage-cost function $\varphi(x,y,p) = -p(x-y) + C(y)$ which is concave in p. This satisfies the conditions of Proposition 3.1, so applying sample-based MPC to (ROSP) will yield the DRO interpretation of Theorem 3.1.

Using the inclusion (3.16) we arrive at the following proposition which presents the optimal policy in closed form.

Proposition 3.2. *Suppose that the distribution $\mathbb{P}$ has compact support. Then the policy which at inventory x and price p sells down to*

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

is optimal for (ROSP).

Proof. Recall that there is a unique continuous function $v : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ satisfying (3.15) and that $x \mapsto v(x,p)$ is concave for each $p \in \mathbb{R}$. Given such v we derive a policy which satisfies (3.16) and therefore solves (ROSP). Define $V : \mathbb{R}_+ \to \mathbb{R}$ by $V(x) \coloneqq \mathbb{E}_{\mathbb{P}}[v(x, \boldsymbol{p})]$ and note that V is

continuous and concave [108, Theorems 7.43, 7.46]. Satisfying (3.16) amounts to solving

$$\underset{y\in\mathbb{R}_+}{\text{maximize}} \quad p(x-y) - C(y) + \beta V(y) \quad \text{subject to} \quad y \in [0,x]. \tag{3.17}$$

Due to the strict concavity of C, (3.17) has a unique solution $y^\star(x,p) \in [0,x]$. Let $y^\star(\infty,p) \coloneqq \lim_{x\to\infty} y^\star(x,p)$. (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 $p \le p^\star(x)$, then $y^\star(\infty,p) \ge x$, and otherwise if $p \ge p^\star(x)$, then $y^\star(\infty,p) \le x$. Comparing $p^\star(x)$ and p characterises whether or not $y^\star(\infty,p)$ needs to be projected onto $[0,x]$ to solve (3.17).

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 $p \le p^\star(x)$, by definition $y^\star(\infty,p) \ge x$, and thus $y^\star(x,p) = x$. In view of (3.15),

$$\begin{aligned} &v(x,p) = -C(x) + \beta V(x) \quad \text{for all } p \le p^\star(x) \\ \implies &-c(x) + \beta \partial V(x) \subseteq \partial_x v(x,p). \end{aligned} \tag{3.18}$$

In the other case,

$$\begin{aligned} v(x,p) = p\big(x - y^\star(\infty,p)\big) - C\big(y^\star(\infty,p)\big) + \beta V\big(y^\star(\infty,p)\big) \quad &\text{for all } p \ge p^\star(x) \\ \implies p \in \partial_x v(x,p). \end{aligned} \tag{3.19}$$

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

$$\psi_x(g,p) \coloneqq \begin{cases} -c(x) + \beta g & \text{if } p \le p^\star(x), \\ p & \text{if } p > p^\star(x); \end{cases} \tag{3.20}$$

the two cases here coinciding with (3.18) and (3.19). Hence, for each $g \in \partial V(x)$,

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

and thus $\mathbb{E}_{\mathbb{P}}[\psi_x(g, \boldsymbol{p})] \in \partial V(x)$ [108, Theorem 7.47]. The expectation of (3.20) is

$$\mathbb{E}_{\mathbb{P}}[\psi_x(g, \boldsymbol{p})] = \mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \leq p^\star(x)\}](\beta g - c(x)) + \mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}], \tag{3.21}$$

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

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

i.e., that $g \mapsto \mathbb{E}_{\mathbb{P}}[\psi_x(g, \boldsymbol{p})]$ 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 $g(x) = \mathbb{E}_{\mathbb{P}}[\psi_x(g(x), \boldsymbol{p})]$. So (3.21) yields

$$\begin{aligned} g(x) &= \mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \leq p^\star(x)\}](\beta g(x) - c(x)) + \mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}] \\ &= \frac{\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^\star(x)\}] - \mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \leq p^\star(x)\}]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \leq p^\star(x)\}]} \in \partial V(x). \end{aligned} \tag{3.22}$$

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 (3.17) is

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

which can be rearranged to

$$\begin{aligned} c(y) &= -(1 - \beta\mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \leq p\}])p + \beta\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p\}] \\ &= \beta\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p\}] - \beta\mathbb{E}_{\mathbb{P}}[p\,\mathbb{1}\{\boldsymbol{p} > p\}] - (1 - \beta)p \\ &= \beta\mathbb{E}_{\mathbb{P}}[(\boldsymbol{p} - p)_+] - (1 - \beta)p. \end{aligned} \tag{3.23}$$

If this is an element of $[c(0), c(x)]$, then the inverse $c^{-1}(\beta\mathbb{E}_{\mathbb{P}}[(\boldsymbol{p} - p)_+] - (1 - \beta)p)$ solves (3.17). Otherwise, due to the concavity of the univariate objective function, first projecting (3.23) onto $[c(0), c(x)]$ solves (3.17). □

The optimal policy of Proposition 3.2 has a natural interpretation: the term $\beta\mathbb{E}_{\mathbb{P}}[(\boldsymbol{p} - p)_+]$ 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 3.2 shows that for each price p problem (ROSP) has an optimal target inventory level $y(\infty, p) \coloneqq \lim_{x\to\infty} y(x, p)$. 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 3.2 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 $p_1, \dots, p_N$, Proposition 3.2 shows that the optimal SDP policy, which we denote by y_{S}, is

$$y_{\mathrm{S}}(x, p; p_1, \dots, p_N) = c^{-1}\Big(\big(\beta \tfrac{1}{N} \textstyle\sum_{i=1}^N (p_i - p)_+ - (1-\beta)p\big)_{[c(0), c(x)]} \Big). \tag{3.24}$$

The optimal MPC policy, which we denote by y_{M}, can then be obtained from Proposition 3.2 by applying it to the point-mass distribution at the sample average $\mu_N = \frac{1}{N}\sum_{i=1}^N p_i$, giving

$$y_{\mathrm{M}}(x, p; p_1, \dots, p_N) = c^{-1}\Big(\big(\beta(\mu_N - p)_+ - (1-\beta)p\big)_{[c(0), c(x)]} \Big). \tag{3.25}$$

In what follows, for simplicity we often suppress the dependence on the price samples $p_1, \dots, p_N$. We denote by $y_{\mathrm{S}}(\infty, p)$ and $y_{\mathrm{M}}(\infty, p)$ the target inventory levels of the policies (3.24) and (3.25), and by $p^\star_{\mathrm{S}}(x)$ and $p^\star_{\mathrm{M}}(x)$ their minimum acceptable prices required for sales. Clearly we have $p^\star_{\mathrm{S}} \le \max\{p_1, \dots, p_N\}$, since this is the highest price the SDP policy considers a possibility. Similar remarks hold for the MPC policy.

Depending on the values of the price samples $p_1, \dots, p_N$, 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 $\frac{1}{N}\sum_{i=1}^N (p_i - p)_+ \ge (\mu_N - p)_+$, and thus $y_{\mathrm{S}}(\infty, p) \ge y_{\mathrm{M}}(\infty, p)$. This reveals that the MPC policy deems it worthwhile to sell at lower prices than the SDP policy does. Indeed, $p^\star_{\mathrm{S}}(x) \ge p^\star_{\mathrm{M}}(x)$ 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.

3.4.1 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 3.3. *The distribution $\mathbb{P}$ has non-negative support, a finite mean, and no atoms.*

Under Assumption 3.3, for every $x, y \in [0, x_1]$ and $\mathbb{P}$-almost every $p \in \mathbb{R}_+$, we have that

$$|p(x-y) - C(y)| \leq px_1 + C(x_1).$$

With $\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}x_1 + C(x_1)]$ finite valued, this shows that Assumption 3.3 for (ROSP) is akin to Assumption 3.2 for (SOC), which enables us to define the out-of-sample performance of policies using functional equations.

Under Assumption 3.3 the out-of-sample performance of the SDP policy (3.24) when starting from initial inventory x, which we denote by $\widetilde{V}_{\mathrm{S}}(x)$, is well defined and satisfies the functional equation

$$\widetilde{V}_{\mathrm{S}}(x) = \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}(x - y_{\mathrm{S}}(x, \boldsymbol{p})) - C(y_{\mathrm{S}}(x, \boldsymbol{p})) + \beta \widetilde{V}_{\mathrm{S}}(y_{\mathrm{S}}(x, \boldsymbol{p}))\Big]. \tag{3.26}$$

Similarly, the out-of-sample performance of the MPC policy (3.25) when starting from initial inventory x, which we denote by $\widetilde{V}_{\mathrm{M}}(x)$, is well defined and satisfies the functional equation

$$\widetilde{V}_{\mathrm{M}}(x) = \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}(x - y_{\mathrm{M}}(x, \boldsymbol{p})) - C(y_{\mathrm{M}}(x, \boldsymbol{p})) + \beta \widetilde{V}_{\mathrm{M}}(y_{\mathrm{M}}(x, \boldsymbol{p}))\Big]. \tag{3.27}$$

We now compare the values of $\widetilde{V}_{\mathrm{S}}(x)$ and $\widetilde{V}_{\mathrm{M}}(x)$ 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 3.2 ([108], Theorem 7.44)**.** *Let $f : \mathbb{R}^m \times \Omega \to \mathbb{R}$ be a random function with expected value $F(x) \coloneqq \mathbb{E}[f(x, \boldsymbol{\omega})]$. 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 $g(\boldsymbol{\omega})$, such that $\mathbb{E}[g(\boldsymbol{\omega})] < \infty$, and for every x, x' in a neighbourhood of $\bar{x}$ and almost every $\omega \in \Omega$, the following inequality holds:*

$$|f(x, \omega) - f(x', \omega)| \leq g(\omega)\|x - x'\|_2.$$

(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 3.3. *Let Assumption 3.3 hold. Then the functions $\widetilde{V}_{\mathrm{S}}$ and $\widetilde{V}_{\mathrm{M}}$ are differentiable.*

Proof. We proceed by verifying that the conditions (i)–(iii) of Lemma 3.2 apply to $\widetilde{V}_{\mathrm{S}}$. Under Assumption 3.3, note that $\widetilde{V}_{\mathrm{S}}$ is well defined and finite valued, satisfying (i). Next we address (iii). Define the function $\widetilde{v}_{\mathrm{S}} : \mathbb{R}_+ \times \mathbb{R}_+^\infty \to \mathbb{R}$ by

$$\widetilde{v}_{\mathrm{S}}(x_1, p_1, p_2, \ldots) \coloneqq \sum_{t=1}^{\infty} \beta^{t-1}\Big(p_t(x_t - x_{t+1}) - C(x_{t+1})\Big), \quad \text{where } x_{t+1} = y_{\mathrm{S}}(x_t, p_t),\ t \in \mathbb{N}. \tag{3.28}$$

If $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots \sim \mathbb{P}$ are independent and identically distributed random price variables, the definition (3.11) 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 $p^\star_{\mathrm{S}}(x_1)$ is the minimum price required for the SDP policy to deem it worthwhile to sell a portion of inventory from x_1, and for any $p \geq p^\star_{\mathrm{S}}(x_1)$, the SDP policy sells down to the target inventory level $y_{\mathrm{S}}(\infty, p) \leq x_1$. It follows that in (3.28) all inventory levels $x_t = x_1$ for $t \leq T$, with T being the first time at which $p_T \geq p^\star_{\mathrm{S}}(x_1)$. The next inventory level $x_{T+1} = y_{\mathrm{S}}(\infty, p_T)$, and along with the prices $p_{T+1}, p_{T+2}, \ldots$, these uniquely determine the remaining inventory levels $x_{T+2}, x_{T+3}, \ldots$ when following the SDP policy. Under Assumption 3.3 the probability distribution $\mathbb{P}$ is atomless which implies that $\mathbb{P}[\boldsymbol{p}_t \neq p^\star_{\mathrm{S}}(x_1)] = 1$. It follows that in (3.28) 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 thus write

$$\frac{\partial \widetilde{v}_{\mathrm{S}}(x_1, p_1, p_2, \ldots)}{\partial x_1} = \beta^{T-1} p_T - \sum_{t=1}^{T-1} \beta^{t-1} c(x_1), \tag{3.29}$$

meeting (iii).

It remains to verify (ii). When attempting to evaluate $\mathbb{E}_{\mathbb{P}^\infty}\big[\frac{\partial}{\partial x_1}\widetilde{v}_{\mathrm{S}}(x_1, \boldsymbol{p}_1, \boldsymbol{p}_2, \ldots)\big]$ using the expression (3.29), the time $\boldsymbol{T}$ and the price $\boldsymbol{p_T}$ are random variables. The expectation of the first term in (3.29) is

$$\mathbb{E}_{\mathbb{P}^\infty}\big[\beta^{\boldsymbol{T}-1}\boldsymbol{p_T}\big] \leq \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p} \,\big|\, \boldsymbol{p} \geq p^\star_{\mathrm{S}}(x_1)\big] \leq \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p} \,\big|\, \boldsymbol{p} \geq \max\{p_1, \ldots, p_N\}\big] < \infty,$$

where the first inequality follows since $\boldsymbol{p_T}$ is the first price greater than $p^\star_{\mathrm{S}}(x_1)$, the second inequality follows since the SDP policy always deems it 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 (3.29) is $\mathbb{E}_{\mathbb{P}^\infty}\left[\sum_{t=1}^{\boldsymbol{T}-1}\beta^{t-1}c(x_1)\right] \leq \frac{1}{1-\beta}c(x_1)$ 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 3.2 applies and the function $\widetilde{V}_{\mathrm{S}}(x)$ is differentiable. Similar reasoning shows that $\widetilde{V}_{\mathrm{M}}(x)$ is differentiable. □

Lemma 3.4. *Let Assumption 3.3 hold. Then the derivatives of the functions $\widetilde{V}_{\mathrm{S}}(x)$ and $\widetilde{V}_{\mathrm{M}}(x)$ are given by*

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

and

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

respectively.

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

$$\widetilde{v}_{\mathrm{S}}(x,p) \coloneqq \begin{cases} -C(x) + \beta\widetilde{V}_{\mathrm{S}}(x) & \text{if } p \leq p^\star_{\mathrm{S}}(x), \\ p(x - y_{\mathrm{S}}(\infty,p)) - C(y_{\mathrm{S}}(\infty,p)) + \beta\widetilde{V}_{\mathrm{S}}(y_{\mathrm{S}}(\infty,p)) & \text{if } p > p^\star_{\mathrm{S}}(x). \end{cases} \tag{3.30}$$

The equation (3.26) shows that $\mathbb{E}_{\mathbb{P}}[\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})] = \widetilde{V}_{\mathrm{S}}(x)$. Recalling that $p^\star_{\mathrm{S}}$ is continuous, for each $p < p^\star_{\mathrm{S}}(x)$ there is a neighbourhood of x values for which the first case of (3.30) holds. Due to Lemma 3.3, the derivative $\frac{d}{dx}\widetilde{V}_{\mathrm{S}}(x)$ exists, and thus

$$\frac{\partial\widetilde{v}_{\mathrm{S}}(x,p)}{\partial x} = -c(x) + \beta\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx}. \tag{3.31}$$

In the other case, for each $p > p^\star_{\mathrm{S}}(x)$ the second case of (3.30) holds and

$$\frac{\partial\widetilde{v}_{\mathrm{S}}(x,p)}{\partial x} = p. \tag{3.32}$$

Similar reasoning to that in the proof of Lemma 3.3 shows $\frac{d}{dx}\widetilde{V}_{\mathrm{S}}(x) = \mathbb{E}_{\mathbb{P}}[\frac{\partial}{\partial x}\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})]$, and with the distribution $\mathbb{P}$ atomless, the event $\boldsymbol{p} = p^\star_{\mathrm{S}}(x)$ is immaterial when evaluating $\mathbb{E}_{\mathbb{P}}[\frac{\partial}{\partial x}\widetilde{v}_{\mathrm{S}}(x,\boldsymbol{p})]$.

Taking the expectation of (3.31) and (3.32) yields

$$\begin{aligned}\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx} &= \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\{\boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}\Big]\Big(\beta\frac{d\widetilde{V}_{\mathrm{S}}(x)}{dx} - c(x)\Big) + \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{S}}(x)\}\Big] \\ &= \frac{\mathbb{E}_{\mathbb{P}}[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{S}}(x)\}] - \mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}]c(x)}{1 - \beta\mathbb{E}_{\mathbb{P}}[\mathbb{1}\{\boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}]}.\end{aligned}$$

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

We use Theorem 3.3 to establish the following proposition which compares the out-of-sample performance of SDP and MPC.

Proposition 3.3. *Let Assumption 3.3 hold. For price samples $p_1, \ldots, p_N$ which determine the sample-based SDP and MPC policies by (3.24) and (3.25), respectively, if the minimum acceptable SDP price $p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)$ is such that*

$$c(x) \ge \beta\mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p} \cdot \mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)\}\Big], \quad \textit{for all } x \in [0, x_1],$$

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, the MPC policy performs at least as well out of sample on (ROSP) as the SDP policy.

Proof. 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 3.2). 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 3.4, we have that

$$\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} = \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\{\boldsymbol{p} \le p^{\star}_{\mathrm{M}}(x)\}\Big]\Big(\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x)\Big) + \mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{M}}(x)\}\Big], \tag{3.33}$$

and similarly

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

Thus, to compare the derivatives it suffices to show that

$$\mathbb{E}_{\mathbb{P}}\Big[\boldsymbol{p}\,\mathbb{1}\{p^{\star}_{\mathrm{M}}(x) < \boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}\Big] \ge \mathbb{E}_{\mathbb{P}}\Big[\mathbb{1}\{p^{\star}_{\mathrm{M}}(x) < \boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}\Big]\Big(\beta\frac{d\widetilde{V}_{\mathrm{M}}(x)}{dx} - c(x)\Big). \tag{3.34}$$

From (3.33) we have

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

where the inequality follows from the statement of the proposition. With $\beta \in (0,1)$ and

$$1-\beta\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{\boldsymbol{p} \le p^{\star}_{\mathrm{M}}(x)\}\big] \ge \mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{p^{\star}_{\mathrm{M}}(x) < \boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}\big],$$

multiplying by $\mathbb{E}_{\mathbb{P}}\big[\mathbb{1}\{p^{\star}_{\mathrm{M}}(x) < \boldsymbol{p} \le p^{\star}_{\mathrm{S}}(x)\}\big]$ yields (3.34), 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 3.3 we conclude that $\widetilde{V}_{\mathrm{M}}(x) \ge \widetilde{V}_{\mathrm{S}}(x)$ for each $x \in [0, x_1]$, as required. □

The condition of Proposition 3.3 has a natural interpretation: $\beta\mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)\}\big]$ 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 that $p_1 \le \cdots \le p_N$ and recall $p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N) \le p_N$. The mapping $p_N \mapsto (p_N - p)_+$ in (3.24) is strictly increasing at the point $p = p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)$ where we have $y_{\mathrm{S}}(\infty, p; p_1, \ldots, p_N) = x$. It follows that $p_N \mapsto p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)$ is strictly increasing, and thus $p_N \mapsto \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{p}\,\mathbb{1}\{\boldsymbol{p} > p^{\star}_{\mathrm{S}}(x; p_1, \ldots, p_N)\}\big]$ is decreasing. If the N price samples are independent and identically distributed according to $\mathbb{P}$, then 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 3.3 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 3.3 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 effect.

If $c(0) = 0$, then Proposition 3.3 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 non-negligible initial inventory levels, we present examples below which show that MPC can still outperform SDP when $c(0) = 0$, due to storage costs accumulated early on.

3.4.2 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 $\boldsymbol{p}_1, \dots, \boldsymbol{p}_N$ used to determine the sample-based policies by (3.24) and (3.25) are independent and identically distributed according to the true distribution $\mathbb{P}$. Expected out-of-sample performance is then given by the expectations of (3.26) and (3.27) under the sampling distribution of $(\boldsymbol{p}_1, \dots, \boldsymbol{p}_N)$.

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 3.4. *Let $\mathbb{P}$ be the* Exponential(1) *distribution, $x_1 = 1$, and $C(x) = \frac{1}{2}x^2$. Then for every 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. To make the sample-dependence explicit, we write $\widetilde{V}_{\mathrm{S}}(x; p_1, \dots, p_N) = \widetilde{V}_{\mathrm{S}}(x)$ and $\widetilde{V}_{\mathrm{M}}(x; p_1, \dots, p_N) = \widetilde{V}_{\mathrm{M}}(x)$. The expected out-of-sample performances of the sample-based policies are then the expectations $\mathbb{E}_{\mathbb{P}^N}\big[\widetilde{V}_{\mathrm{S}}(x; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N)\big]$ and $\mathbb{E}_{\mathbb{P}^N}\big[\widetilde{V}_{\mathrm{M}}(x; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N)\big]$.

First consider the term $\mathbb{E}_{\mathbb{P}^N}\big[\widetilde{V}_{\mathrm{S}}(1; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N)\big]$. With $\widetilde{V}_{\mathrm{S}}(0; p_1, \dots, p_N) = 0$, we write $\widetilde{V}_{\mathrm{S}}(1; p_1, \dots, p_N) = \int_0^1 \frac{d}{dx}\widetilde{V}_{\mathrm{S}}(x; p_1, \dots, p_N)\, dx$ using the expression for the derivative given by Lemma 3.4. Hence,

$$\mathbb{E}_{\mathbb{P}^N}\big[\widetilde{V}_{\mathrm{S}}(1; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N)\big] = \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\dots,\boldsymbol{p}_N)}^{\infty} p f(p)\, dp - xF(p^\star_{\mathrm{S}}(x; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N))}{1 - \beta F(p^\star_{\mathrm{S}}(x; \boldsymbol{p}_1, \dots, \boldsymbol{p}_N))}\, dx\right], \quad (3.35)$$

where F and f denote the cumulative distribution function and probability density function of

the Exponential(1) distribution, respectively. The negative term in (3.35) is

$$\mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{-xF(p_{\mathrm{S}}^\star(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N))}{1-\beta F(p_{\mathrm{S}}^\star(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N))}\,dx\right] = \int_0^\infty \cdots \int_0^\infty \int_0^1 \frac{-xF(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}{1-\beta F(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}\,dx\, f(p_N)\,dp_N \cdots f(p_1)\,dp_1. \quad (3.36)$$

Fix $p_1,\ldots,p_{N-1}$ and consider the inner-most integral in (3.36) when p_N is large. Following (3.24), for each $\alpha \in (0,1)$ and all $\beta \in [\alpha,1)$, $x \in [0,1]$, the value $p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N)$ is the p which solves the equation $\beta\frac{1}{N}(p_N-p)_+ + (1-\beta)p = x$. Thus

$$\frac{-xF(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}{1-\beta F(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))} \to \frac{-x(1-\exp(Nx-p_N))}{1-(1-\exp(Nx-p_N))}$$

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

$$\begin{aligned}\lim_{\beta\to 1}\int_0^1 \frac{-xF(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}{1-\beta F(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}\,dx &= \int_0^1 \frac{-x(1-\exp(Nx-p_N))}{1-(1-\exp(Nx-p_N))}\,dx \\ &= \frac{1}{2} + \left(\frac{1}{N}+\frac{1}{N^2}\right)\exp(p_N-N) - \frac{1}{N^2}\exp(p_N).\end{aligned}$$

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

$$\int_0^1 \frac{-xF(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}{1-\beta F(p_{\mathrm{S}}^\star(x;p_1,\ldots,p_N))}\,dx < \frac{1}{2} + \left(\frac{1}{N}+\frac{1}{N^2}\right)\exp(p_N-N) - \frac{1}{N^2}\exp(p_N) + \varepsilon. \quad (3.37)$$

But

$$\begin{aligned}&\lim_{p_N\to\infty}\left(\frac{1}{2} + \left(\frac{1}{N}+\frac{1}{N^2}\right)\exp(p_N-N) - \frac{1}{N^2}\exp(p_N) + \varepsilon\right)\exp(-p_N) \\ &= \left(\frac{1}{N}+\frac{1}{N^2}\right)\exp(-N) - \frac{1}{N^2} < 0,\end{aligned}$$

and for $0 < a < b$ the integral

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

can be made arbitrarily negative by increasing b. Using the upper bound (3.37), it follows that

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

can be made arbitrarily negative by increasing β towards 1, regardless of the values of $p_1, \ldots, p_{N-1}$. We therefore conclude (3.36) is unbounded below as $\beta \to 1$.

Recall that $p^\star_{\mathrm{S}}(x; p_1, \ldots, p_N) \le \max\{p_1, \ldots, p_N\}$. The positive term in (3.35) is

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N)}^\infty p f(p)\,dp}{1-\beta F(p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N))}\,dx\right] &\le \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 \frac{\int_{p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N)}^\infty p f(p)\,dp}{1-F(p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N))}\,dx\right] \\
&= \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 1 + p^\star_{\mathrm{S}}(x;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N)\,dx\right] \\
&\le \mathbb{E}_{\mathbb{P}^N}\left[\int_0^1 1 + \max\{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N\}\,dx\right] \\
&= \mathbb{E}_{\mathbb{P}^N}\Big[1 + \max\{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N\}\Big].
\end{aligned} \tag{3.38}$$

Here $\mathbb{E}_{\mathbb{P}^N}\big[\max\{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N\}\big]$ is the expected value of the N-th order statistic of N random samples from the Exponential(1) distribution, which is finite; see, e.g., [30, Section 4.2]. Hence, (3.38) is bounded above for all $\beta \in (0,1)$. We have already shown that (3.36) is unbounded below as $\beta \to 1$. Thus, the sum (3.35) is unbounded below as $\beta \to 1$, regardless of the choice of N.

Consider now $\mathbb{E}_{\mathbb{P}^N}\big[\widetilde{V}_{\mathrm{M}}(1;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N)\big]$. For simplicity, we set $N=2$, although the result can be shown to hold for $N \ge 2$ in general. As for (3.35), we have that

$$\mathbb{E}_{\mathbb{P}^2}\big[\widetilde{V}_{\mathrm{M}}(1;\boldsymbol{p}_1,\boldsymbol{p}_2)\big] = \mathbb{E}_{\mathbb{P}^2}\left[\int_0^1 \frac{\int_{p^\star_{\mathrm{M}}(x;\boldsymbol{p}_1,\boldsymbol{p}_2)}^\infty p f(p)\,dp - xF(p^\star_{\mathrm{M}}(x;\boldsymbol{p}_1,\boldsymbol{p}_2))}{1-\beta F(p^\star_{\mathrm{M}}(x;\boldsymbol{p}_1,\boldsymbol{p}_2))}\,dx\right], \tag{3.39}$$

which has the negative term

$$\mathbb{E}_{\mathbb{P}^2}\left[\int_0^1 \frac{-xF(p^\star_{\mathrm{M}}(x;\boldsymbol{p}_1,\boldsymbol{p}_2))}{1-\beta F(p^\star_{\mathrm{M}}(x;\boldsymbol{p}_1,\boldsymbol{p}_2))}\,dx\right] = \int_0^\infty\int_0^\infty\int_0^1 \frac{-xF(p^\star_{\mathrm{M}}(x;p_1,p_2))}{1-\beta F(p^\star_{\mathrm{M}}(x;p_1,p_2))}\,dx\,f(p_2)\,dp_2\,f(p_1)\,dp_1. \tag{3.40}$$

The iterated integral (3.40) can be divided into ranges based on the value of the sample average $\mu_2 = \frac{1}{2}(p_1+p_2)$. Following (3.25), depending on the value of μ_2, either $p^\star_{\mathrm{M}}(x;p_1,p_2) = \beta\mu_2 - x$ or $p^\star_{\mathrm{M}}(x;p_1,p_2) = -x/(1-\beta)$. Since $\mu_2 - x \ge \beta\mu_2 - x \ge -x/(1-\beta)$, we have that

$$\begin{aligned}
&\int_0^1 \frac{-xF(p^\star_{\mathrm{M}}(x;p_1,p_2))}{1-\beta F(p^\star_{\mathrm{M}}(x;p_1,p_2))}\,dx \\
\ge\ &\int_0^1 \frac{-xF(\mu_2-x)}{1-F(\mu_2-x)}\,dx \\
=\ &\int_0^{\min\{\mu_2,1\}} \frac{-x(1-\exp(x-\mu_2))}{1-(1-\exp(x-\mu_2))}\,dx \\
=\ &\frac{1}{2}(\min\{\mu_2,1\})^2 + (1+\min\{\mu_2,1\})\exp(\mu_2-\min\{\mu_2,1\}) - \exp(\mu_2).
\end{aligned}$$

Thus, using this last term, (3.40) is bounded below by

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

for all $\beta \in (0,1)$. Similar reasoning as for (3.38) shows that the positive term in (3.39) is bounded above. We therefore conclude the sum (3.39) is bounded as $\beta \to 1$. □

Proposition 3.4 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 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. (It is important to note this does not conflict with the epi-consistency results of the previous Chapter 2, since those are for a fixed discount factor.) Moreover, in real-world 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` [19]. 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 $T = 10^3$.[2] With $x_1 = 1$, $\beta = 0.99$, and $C(x) = \frac{1}{2}x^2$, for a given sample size N we repeat:

0. Sample $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N \sim \mathbb{P}$ to determine a policy y using (3.24) for SDP and (3.25) for MPC.

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

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

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 10^5 realisations, finding this sufficient to ensure accurate results. (In Figures 3.1–3.5 the standard error ranges are smaller than the markers so are not shown.)

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

3.4.3 Skewed Price Distributions

Let Triangular(a, m, b) denote the triangular distribution with lower limit a, mode m, and upper limit b. Figures 3.1 and 3.2 present the expected out-of-sample performance of SDP and MPC for left- and right-skewed triangular price distributions with mean 1 and variance $1/8$.

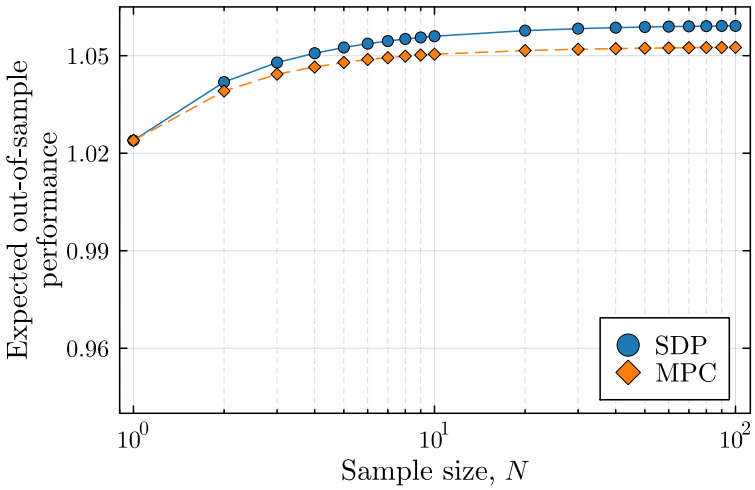

Figure 3.1: **Performance for Triangular$(0, 3/2, 3/2)$ Distributed Price**

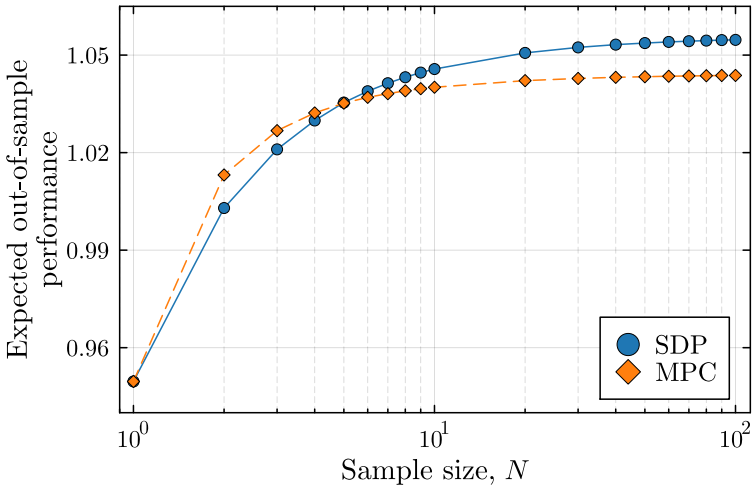

Figure 3.2: **Performance for Triangular$(1/2, 1/2, 2)$ Distributed Price**

Figure 3.1 shows SDP outperforming MPC for all sample sizes when the true price distribution is left skewed. In contrast, Figure 3.2 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 3.3 presents the expected out-of-sample performance of SDP and MPC for a right-skewed triangular price distribution with mean 1 and variance $1/2$.

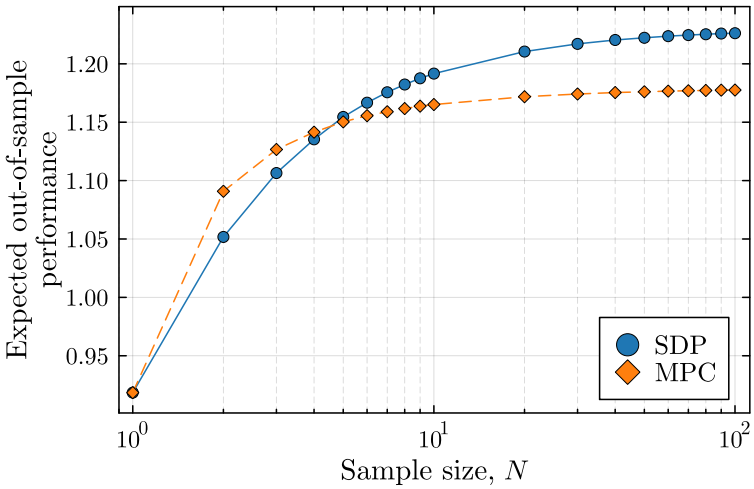

Figure 3.3: **Performance for Triangular(0, 0, 3) Distributed Price**

Figure 3.3 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 3.2. This is the same range as that in Figure 3.2, but here the relative amount by which MPC outperforms SDP is increased.

3.4.4 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 3.4 and 3.5 present the expected out-of-sample performance of SDP and MPC for exponential and lognormal price distributions with mean 1.

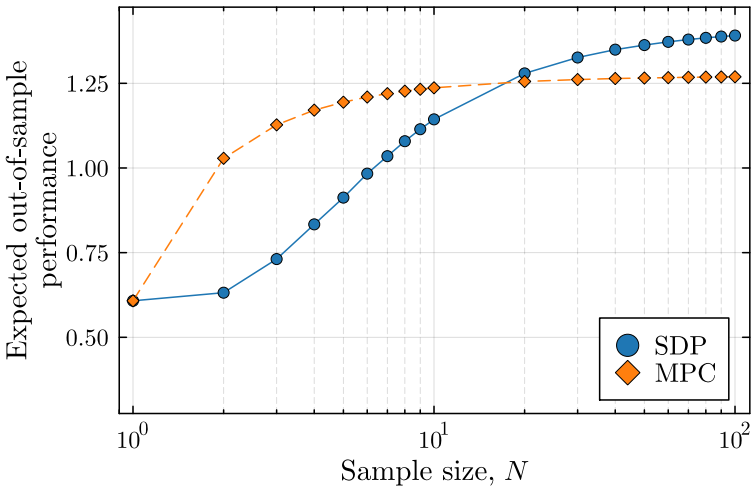

Figure 3.4: **Performance for Exponential(1) Distributed Price**

Figure 3.4 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 3.1–3.3. 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.

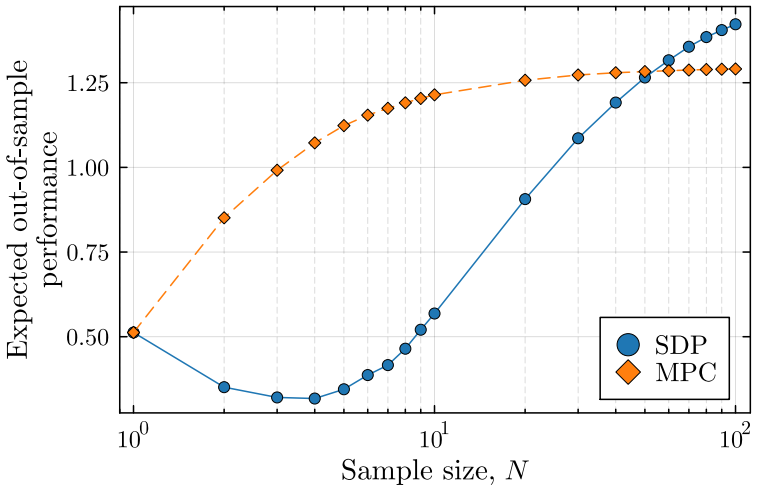

Figure 3.5: **Performance for LogNormal($-1/2, 1$) Distributed Price**

Figure 3.5 shows MPC outperforming SDP for sample sizes $N \leq 50$ when the true price distribution is right skewed and has a heavier right-hand tail than the exponential distribution used in Figure 3.4. The poor performance of SDP exhibited here suggests a result similar to that of Proposition 3.4 for exponential distributions may also hold for lognormal distributions. Considering lognormal distributions have more weight in their tails than exponential distributions, this would not be surprising.

For a sample of size $N = 2$, Figure 3.6 presents the difference in the out-of-sample performance of the SDP and MPC policies.

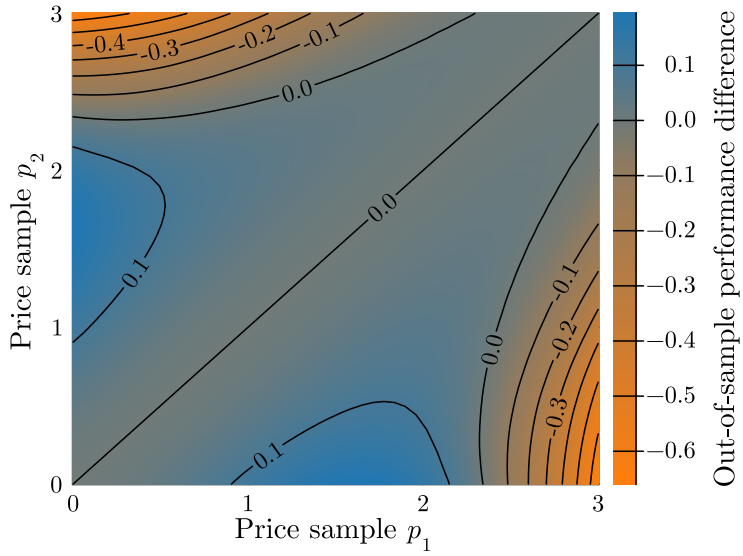

Figure 3.6: **Performance Difference for LogNormal($-1/2, 1$) Distributed Price.** $N = 2$ samples. Positive values indicate SDP outperforms MPC and negative values indicate the reverse.

Figure 3.6 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 3.3.

3.5 Discussion

In this chapter 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 DRO 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 [35], and that for which the optimistic interpretation holds features constraint uncertainty, as in [90]. 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 [35], 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 problem with the discount factor $\beta = 1 - \gamma$.

Chapter 4

Supply Chain Management in the New Zealand Dairy Industry

In the previous chapter we investigated the performance of sample-based approximations of stochastic optimal control problems. Owing to the practical necessity of using small sample sizes for industrial-scale applications, we showed that model predictive control (MPC) can be a better choice than stochastic dynamic programming (SDP) when problems feature objective-function uncertainty. Conversely, we showed that SDP can be a better choice than MPC when problems feature constraint uncertainty. This objective-function uncertainty is a characteristic of resource allocation under price uncertainty, whereas constraint uncertainty is a characteristic of resource-supply uncertainty.

Building on our understanding of what problem characteristics drive the performance differences of SDP and MPC, in this chapter we study a supply chain–management problem arising in the New Zealand dairy industry. Aspects of the problem include price uncertainty and resource-supply uncertainty. We aim to develop a well-suited data-driven approximation scheme. Assuming that the underlying stochasticity does not change over time, we simulate the evolution of international dairy commodity prices by fitting an autoregressive process to real-world data. This allows us to test the performance of different approximation schemes, and we compare to an industry decision-making policy.

This chapter is based on joint work with *Andrew B. Philpott & Andrew J. Mason.*

The New Zealand dairy industry we study is a *cooperative*: a central entity that collects raw milk from farmers to process and sell on international dairy commodity markets. In return for their milk farmers receive a payout from the total operating profit of the cooperative at the end of the dairy season [29]. As the dairy season progresses, forecasts of this payout are provided to farmers ahead of time. Such forecasts are derived from models of international commodity prices and the production and sales process itself, which incorporates many sources of uncertainty [36, 39]. Whatever model is employed may systematically under- or over-estimate the payout. The payout milk price is a New Zealand dairy farmer's livelihood, and payout disappointment can negatively impact farmers' lives [56]. Moreover, the forecasting process also informs the operational decision making of the cooperative, and poor estimates may result in production and sales decisions which yield suboptimal profits. Therefore, sound production and sales planning models are beneficial to the continued success of the New Zealand dairy industry.

Guan & Philpott [59, 60] consider the effects of uncertain milk supply on operational decision making within the New Zealand dairy industry. They develop a stochastic programming model in which international commodity prices are not stochastic, but determined based on the amount of product sold via the solution of an embedded equilibrium problem with known price–demand curves. However, as identified in [36], due to the coincidence of market deregulation, auction platform introductions, and consumption changes, international dairy commodities have exhibited a marked increase in price volatility since 2008; see Figure 4.1.

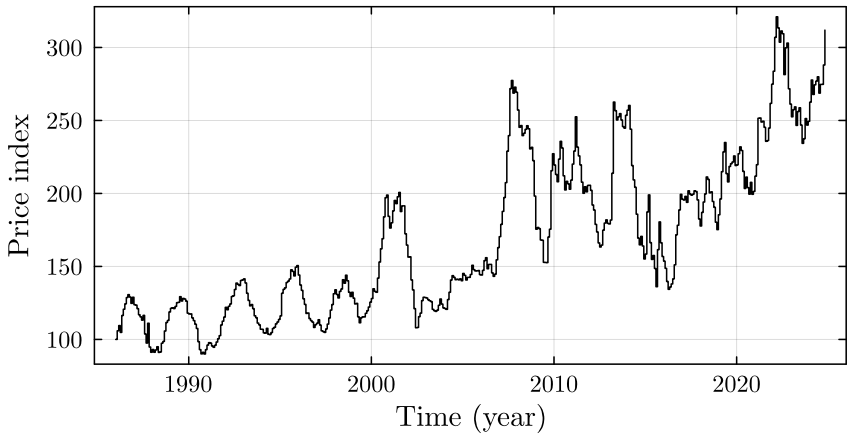

Figure 4.1: **Historical Dairy Commodity Price Index.** Jan. 1986–Nov. 2024.[1]

Ignoring such volatility when planning poses a significant risk. To address this, [58] develops a multistage stochastic programming model incorporating price uncertainty. In this chapter we show that the disappointing performance of the resulting decision-making policy against an

[1]Data retrieved from [11].

equivalent MPC policy can be explained by the results of the previous chapter. While a further study shows that the aforementioned model has significant value against certain proprietary policies [111], we show that this value is simply due to accurate price forecasting. Consequently, this benefit is also exhibited by an MPC policy following reasonable price forecasts.

The remainder of this chapter is organised as follows. In Section 4.1 we characterise the general production and sales optimization problem faced by dairy cooperatives. Using real data, we model market prices as a multiplicative lag-one vector-autoregressive process and formulate the optimization problem as a multistage stochastic program that can be solved with variants of the stochastic dual dynamic programming algorithm. In Section 4.2 we observe that the resulting multistage stochastic program has similarities to the problem (ROSP) studied in Chapters 3, and hence we formulate the MPC version of the production and sales optimization problem. Finally, in Section 4.3 we compare the performance of the MPC policy to a proprietary policy via simulations based on real data.

4.1 Multistage Stochastic Programming Model

The decision-making problem faced by New Zealand dairy cooperatives is characterised by a supply chain manager coordinating the allocation of periodic resource inflows (raw milk) into different production processes for commodities (dairy products). Each process yields different mixtures of products, as diagrammatised by Figure 4.2.

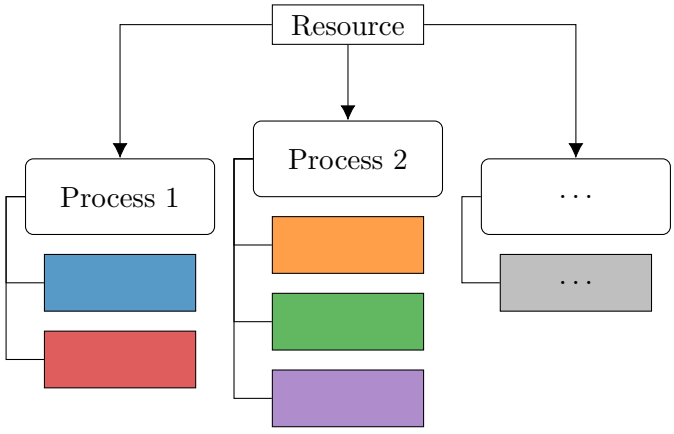

Figure 4.2: **Production Diagram**

Product can then be sold on the market immediately, or stored for sale at a later date, subject to a storage cost. The market prices are stochastic, and the supply chain manager wishes to make production and sales decisions to maximize their expected future profit.

At the start of the first decision-making period inventory is initialised based on that remaining from the end of a previous season or decision epoch. Beyond the final decision-making period, inventory is viewed as having no value. For a sufficiently long planning horizon this is a realistic simplification, as we apply the model in a rolling-horizon fashion such that the planning horizon always extends beyond the decision at hand.

Each commodity has different storage requirements. We assume that each commodity can be stored at a per-period storage cost which is a linear function of the amount stored. We also assume that each storage warehouse has a finite storage capacity. Commodities which perish can be treated in this framework by ensuring storage costs are sufficiently large, or that there is no associated storage capacity. Some products share common storage facilities, and shared capacity constraints can be enforced through restrictions to the set of feasible inventories.

We model the market price for each commodity as a correlated random variable that evolves over time. We neglect modelling price elasticity. We also neglect modelling futures contracts that may be offered by some large consumers, although our eventual model could be easily extended to account for these by enforcing shortage costs (with suitable linkage constraints introduced).

The supply chain manager operates a number of different processes which each take raw resource and transform it into a mixture of different commodities in different ratios (see Figure 4.2). Hence, each production process has an operating cost and a maximum processing capacity. This amounts to a per-unit production cost and a restriction of the feasible production decisions. We neglect modelling any sort of explicit transportation costs, instead assuming that these can be reflected as operating costs. As the lesser source of stochasticity in the model, the raw supply of resources available to be processed into different products is modelled deterministically for simplicity.

For m products and a planning horizon T, we define the following:

Indices

$t \in [T] \coloneqq$ decision-making period.

Parameters

$c \in \mathbb{R}^m_+ \coloneqq$ per-unit per-period storage cost,

$h \in \mathbb{R}^m_+ \coloneqq$ per-unit production cost,

$\beta \in (0, 1) \coloneqq$ per-period discount factor.

State Variables

$x \in \mathbb{R}^m_+ :=$ inventory of product,

$p_1, \ldots, p_t \in \mathbb{R}^m_+ :=$ historical per-unit market prices observed in periods 1 to t.

Random Variables

$\boldsymbol{p}_{t+1} \in \mathbb{R}^m_+ :=$ random per-unit market price in period $t+1$, given historical prices $p_1, \ldots, p_t$.

Decision Variables

$u \in \mathbb{R}^m_+ :=$ units of product to sell from inventory,

$w \in \mathbb{R}^m_+ :=$ units of product to produce and add to inventory.

Feasible Sets

$\mathcal{X} \subseteq \mathbb{R}^m_+ :=$ inventory that the manager has capacity to store between periods,

$\mathcal{W} \subseteq \mathbb{R}^m_+ :=$ product that the manager has capacity to produce between periods.

Value-Function Recursion

Let the functions $v_t, t \in [T]$ satisfy

$$
\begin{aligned}
&v_t(x, p_1, \ldots, p_t) && \text{(SCM)}\\
&\quad = \sup_{u,w \in \mathbb{R}^m_+} \Big\{ p_t^\intercal u - c^\intercal (x-u) - h^\intercal w + \beta \mathbb{E}\Big[v_{t+1}(x-u+w, p_1, \ldots, p_t, \boldsymbol{p}_{t+1}) \,\Big|\, p_1, \ldots, p_t \Big] \Big\},\\
&\qquad \text{s.t.} \quad u \in [0, x],\\
&\qquad\qquad\;\; x - u + w \in \mathcal{X},\\
&\qquad\qquad\;\; w \in \mathcal{W},
\end{aligned}
$$

where $v_{T+1} := 0$. Note that these functions are well defined and real valued for reasonable price processes, if, for instance, $\mathcal{X}$ and $\mathcal{W}$ are nonempty and compact. To this end, throughout the chapter we work with $\mathcal{X}$ and $\mathcal{W}$ which are compact convex polytopes.

At period t, for a given inventory x and historical market prices $p_1, \ldots, p_t$, the optimal production and sales decision is the pair (u, w) attaining the supremum in (SCM). We treat the decisions to be made in periods t and $t+1$ as those in consecutive months of a dairy season and vary the planning horizon T as required. We set the discount factor $\beta = \sqrt[12]{0.9}$, corresponding to a yearly discount factor of 0.9. For the other parameters in the model we use the values of [111]. For simplicity we also set the monthly supply of raw milk to be a constant over the dairy season, using the monthly average.

The supply chain–management problem (SCM) is an instance of Example 1.1 reformulated as finding solutions to SDP functional equations. Whether or not solution functions satisfy the principle of optimality depends on the price process. We elaborate further on a specific case in the next subsection.

4.1.1 Stochastic Price Process

The New Zealand dairy industry produces a wide range of products, but for simplicity we only model the stochastic process behind the prices of: Anhydrous Milk Fat (AMF), Butter (BUT), Butter Milk Powder (BMP), Skim Milk Powder (SMP), and Whole Milk Powder (WMP). We also include a number of additional products in later optimizations, but hold their market prices constant at proprietary values.

Figure 4.3 graphs fourteen years of Global Dairy Trade (GDT) prices for the five products. The GDT price is the average per-unit price of each product sold at international commodity auctions in each month. We use this as a proxy for international market spot prices.

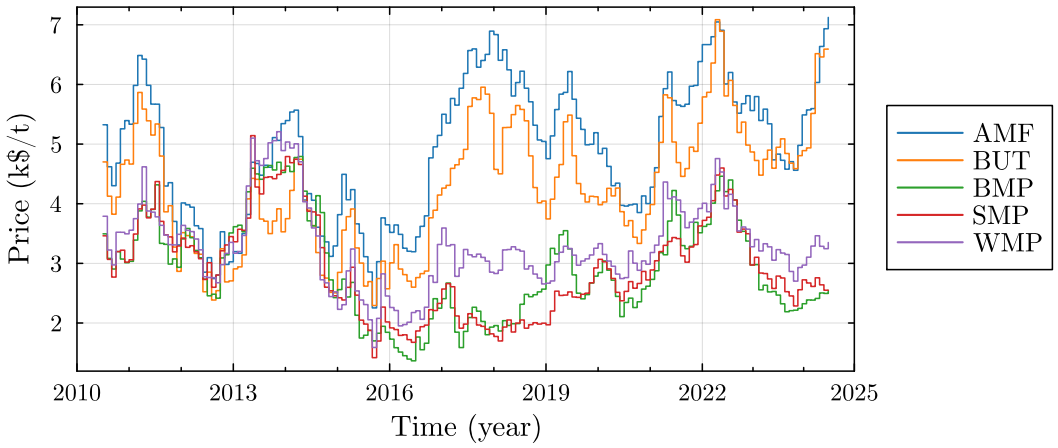

Figure 4.3: **Historical GDT Prices.** Monthly observations; June 2010–May 2024.[2,3]

Figure 4.3 shows that the GDT prices of each product are correlated. In particular, the prices of the fat products (AMF and BUT) and the milk powders (BMP, SMP, and WMP) are coupled throughout the fourteen-year period. From Figure 4.3 it is also evident that the GDT prices of each product are serially correlated. Figure 4.4 graphs the lagged partial autocorrelations of each time series in Figure 4.3.

[3]Data retrieved from [39].
[3]We use this legend colour scheme for the five dairy products throughout the chapter.

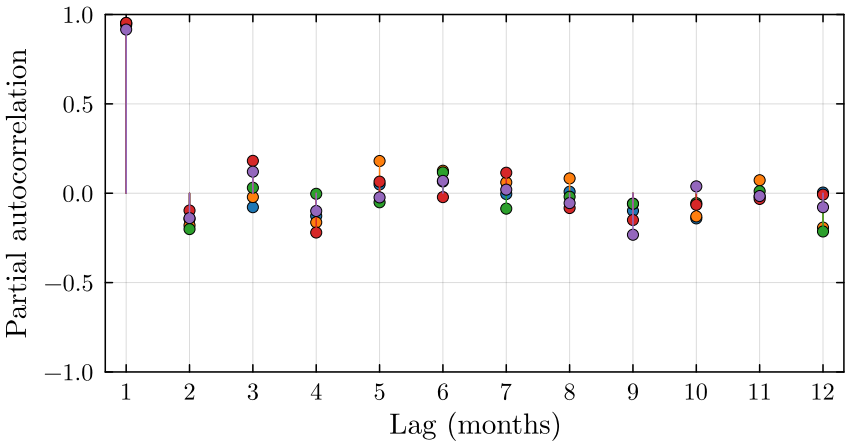

Figure 4.4: **GDT-Price Partial Autocorrelation**

Figure 4.4 shows that future GDT prices are correlated most strongly with their current values, and that this correlation is positive.

Instead of modelling the random prices directly, we model their natural logarithms. This captures the multiplicative effect of global supply chain shortages and surpluses on international market prices. Between each month we view the current market price vector $p_t \in \mathbb{R}^m$ as evolving randomly through the vector equation

$$\log(\boldsymbol{p}_{t+1}) = \mu + A \cdot \log(p_t) + \boldsymbol{\xi}, \tag{4.1}$$

where $\mu \in \mathbb{R}^m$ is a vector, $A \in \mathbb{R}^{m \times m}$ is a matrix, and $\boldsymbol{\xi} \in \mathbb{R}^m$ is a random vector. That is, we suppose that the market prices follow a multiplicative lag-one vector-autoregressive process.

Although the New Zealand dairy industry exports a massive amount of product, this is still less than 5 % of the global dairy supply [29], so we neglect modelling price elasticity as influenced by our own decisions. As further evidence for the validity of this simplification, due to the seasonality of milk production by dairy cows, the New Zealand dairy industry operates on repeating, yearly, cycles, but such an effect is not present in the historical price data, being in neither Figure 4.3 nor 4.4.

Given the sequence of historical price observations in Figure 4.3, each line in the vector equation (4.1) yields an ordinary linear-regression problem. By capturing correlation between each regression's error terms, sometimes there is an efficiency gain possible from estimating a system of regressions jointly using approaches such as seemingly unrelated regression. However, in the present context the regressor in each line of the vector equation (4.1), being $\log(p_t)$, is the same, which means that the resulting seemingly unrelated regression estimate is identical to that

returned by applying ordinary least squares to each line separately [32, Section 9.8]. Table 4.1 shows the resulting A matrix, exhibiting the coupled relationships of the fat products and the milk-powder products.

Table 4.1: **Ordinary Least-Squares Fit A Matrix**

Future log price	Current log price				
	AMF	BUT	BMP	SMP	WMP
AMF	0.88	0.02	0.03	−0.15	0.13
BUT	0.18	0.78	0.00	−0.02	0.00
BMP	−0.03	−0.05	0.83	−0.03	0.26
SMP	−0.10	0.02	0.13	0.72	0.17
WMP	0.07	−0.03	0.09	−0.01	0.81
Eigenvalue	0.61	0.67	0.83	0.95	0.97

With μ and A values in hand, each pair of successive price observations (p_t, p_{t+1}) from the data in Figure 4.3 defines an empirical process noise $\xi_t := \log(p_{t+1}) - \mu - A\log(p_t)$. We treat the resulting values $\{\xi_1, \ldots, \xi_N\}$ as samples from the random variable $\boldsymbol{\xi}$ and hence discretise its underlying probability distribution. Figure 4.5 graphs simulations of random price trajectories following this discretised model.

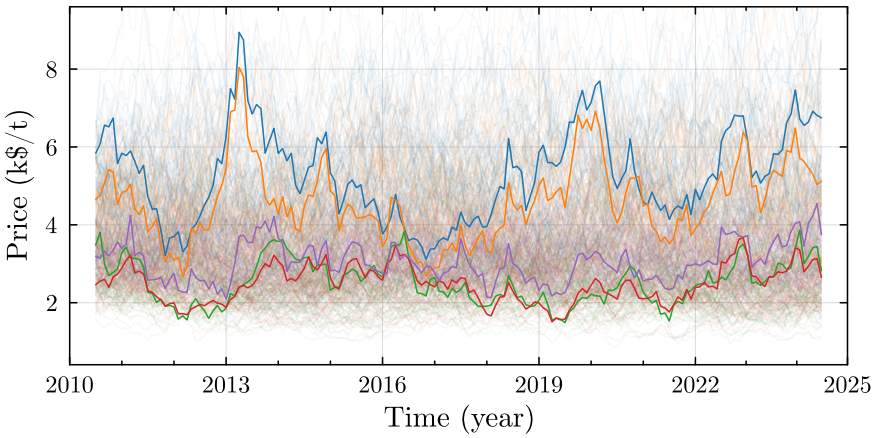

Figure 4.5: **Simulated Prices.** Random initialisations.

Following this discretised model and using ℓ to represent the log-price variable, the value-function

recursion (SCM) becomes

$$v_t(x,\ell) = \sup_{u,w\in\mathbb{R}^m_+} \Big\{ \exp(\ell)^\intercal u - c^\intercal(x-u) - h^\intercal w + \beta \tfrac{1}{N} \textstyle\sum_{i=1}^N v_{t+1}\big(x-u+w, \mu + A\ell + \xi_i\big) \Big\},$$
$$\text{s.t.} \quad u \in [0,x], \tag{4.2}$$
$$x - u + w \in \mathcal{X},$$
$$w \in \mathcal{W}.$$

Observe that the resulting value functions $v_1, \ldots, v_T$ are saddle functions, and with $\mathcal{X}$ and $\mathcal{W}$ being compact convex polytopes, $v_1, \ldots, v_T$ can be solved for using the "objective state" variant of the classical stochastic dual dynamic programming algorithm [35] (see also [36, Chapter 4]).

Regarding the principle of optimality, consider the fact that all of the eigenvalues in Table 4.1 have magnitudes strictly less than one. Viewing equation (4.1) as a vector-linear recurrence with constant coefficients, we can argue that for an initial log price ℓ_1 and any sequence of realisations from the noise samples $\xi_1, \ldots, \xi_N$, the resulting sequence of future log prices $\ell_2, \ell_3, \ldots$ is bounded, regardless of the length of the planning horizon T. Since both $\mathcal{X}$ and $\mathcal{W}$ are also bounded, it follows that the stage-cost function is also bounded. Consequently, the unique continuous value functions $v_1, \ldots, v_T$ satisfy the principle of optimality relative to the approximating control problem of Example 1.1; see, e.g., [113, Theorems 9.2 and 9.6].

4.2 Model Predictive Control

With a suitable translation into a state–noise formulation, the Bellman operator associated with the value-function recursion (4.2) is concavity preserving. In the previous chapter we discovered that MPC can have an expected out-of-sample performance benefit over SDP due to the presence of this concavity preservation. (For details, see the definitions and examples of Section 3.2.) Motivated by this, we now construct the MPC version of (4.2). Replacing each of the ξ_i terms with the mean $\frac{1}{N}\sum_{i=1}^N \xi_i$ and noting that this is equal to 0 due to the ordinary least-squares fit,

we obtain the value-function recursion

$$v_t(x,\ell) = \sup_{u,w\in\mathbb{R}^m_+} \Big\{ \exp(\ell)^\intercal u - c^\intercal(x-u) - h^\intercal w + \beta v_{t+1}(x-u+w, \mu + A\ell) \Big\}, \qquad (4.3)$$

$$\text{s.t.} \quad u \in [0, x],$$

$$x - u + w \in \mathcal{X},$$

$$w \in \mathcal{W}.$$

Figure 4.6 graphs some examples of the price trajectories within this recursion.

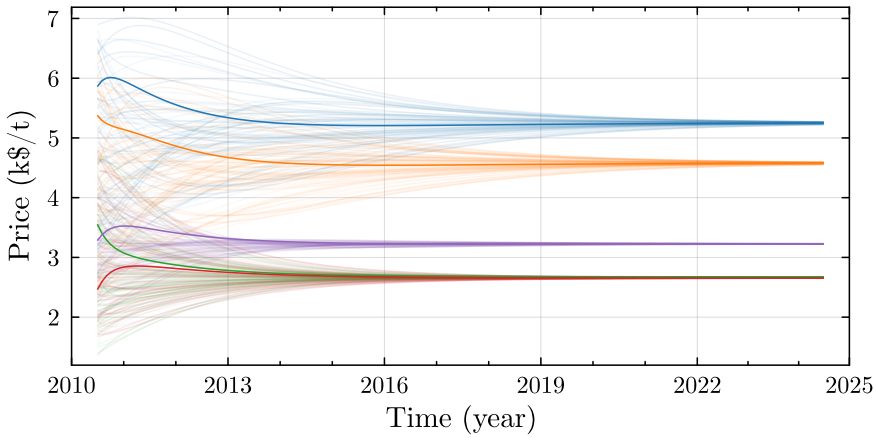

Figure 4.6: **Simulated Prices (Zero Process Noise).** Random initialisations.

We formulate the problem associated with the MPC value-function recursion (4.3) in `Julia` [19] using `JuMP.jl` [41] and solve with `Gurobi` [62]. The problem is deterministic, and with $\mathcal{X}$ and $\mathcal{W}$ being compact convex polytopes, it amounts to a linear program which can be solved quickly. Figures 4.7 and 4.8 graph decision trajectories following the resulting policy.

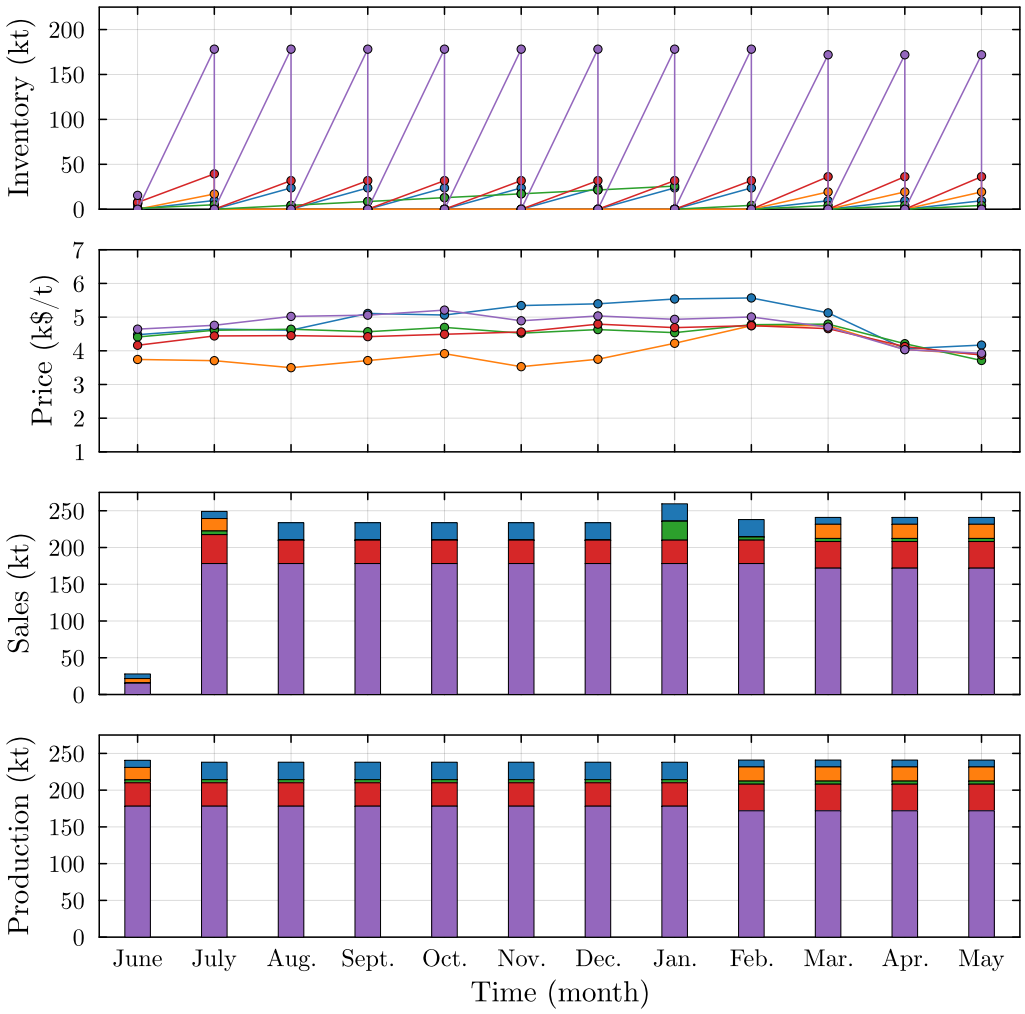

Figure 4.7: **High Price Sequence Decision Trajectory.** Each month the inventory pane shows the initial inventory, an instantaneous drop due to sales at the current price, and an increase over the remainder of the month due to production.

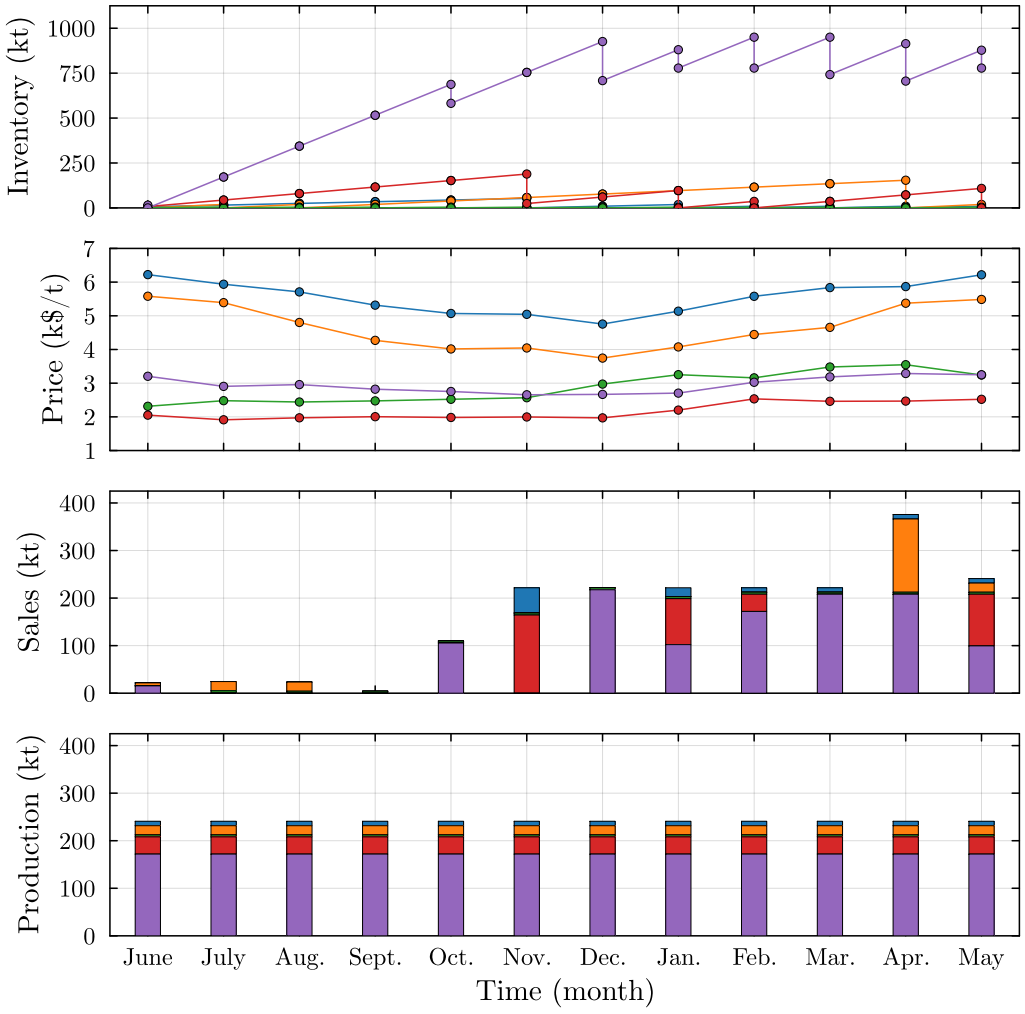

Figure 4.8: **Low Price Sequence Decision Trajectory.** Each month the inventory pane shows the initial inventory, an instantaneous drop due to sales at the current price, and an increase over the remainder of the month due to production.

Figure 4.7 shows the MPC policy immediately selling all available WMP from inventory when the WMP market price is high. Conversely, when the WMP market price is low, Figure 4.8 shows the MPC policy accumulating WMP until a certain threshold is reached and then selling to maintain this level. For the other products Figure 4.8 shows the MPC policy selling all available inventory as the market prices creep up in the latter half of the dairy season. This selling behaviour is reminiscent of the target inventory levels found to be optimal in Section 3.4.

Both Figures 4.7 and 4.8 show that the production decisions are relatively stable throughout the season and dominated by WMP. This is unsurprising as WMP makes up the majority of New Zealand dairy exports, so the investments made by cooperatives into production capacity would reflect this.

4.3 Policy Performance

We now compare through simulation the performance of the MPC policy developed in Section 4.2 to that of a proprietary policy previously used by a New Zealand dairy cooperative. We simulate price sequences using the discrete autoregressive process of Section 4.1. To assess out-of-sample performance, we use a sample-based version of MPC where the autoregression parameters are fitted to simulated historical price sequences. To provide an upper bound on the best possible performance, we also test a clairvoyant policy which has perfect foresight of the future. For each simulation we realise a profit following each policy. The averages of these statistics provide estimates of expected out-of-sample profit. We set the total time horizon of each simulation to one hundred years which provides a reasonable approximation of the infinite-horizon problem with our yearly discount factor of 0.9. Figure 4.9 graphs the resulting expected profits.

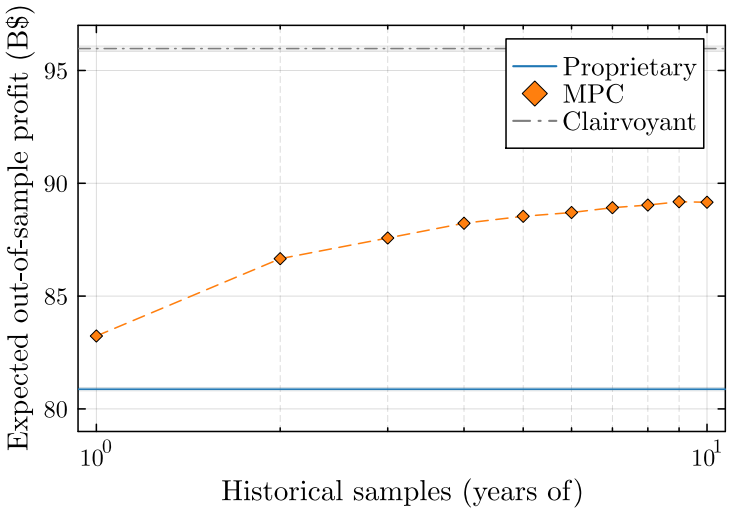

Figure 4.9: **MPC Profit Benchmark.** Standard-error ranges within the marker size.

Figure 4.9 shows that, on average, sample-based MPC outperforms the proprietary policy, even with only a single year of historical data. With ten years of historical data, MPC is able to achieve out-of-sample performance levels near its asymptotic limit. In view of the consistency results of Chapter 2, this is unsurprising; see in particular Subsection 2.4.3.

Given that only ten years of historical data are sufficient to attain MPC out-of-sample performance levels near its asymptotic limit, henceforth we work with the discrete autoregressive process of Section 4.1 directly, i.e., without sampling. Figure 4.10 graphs the resulting profit distributions of each policy.

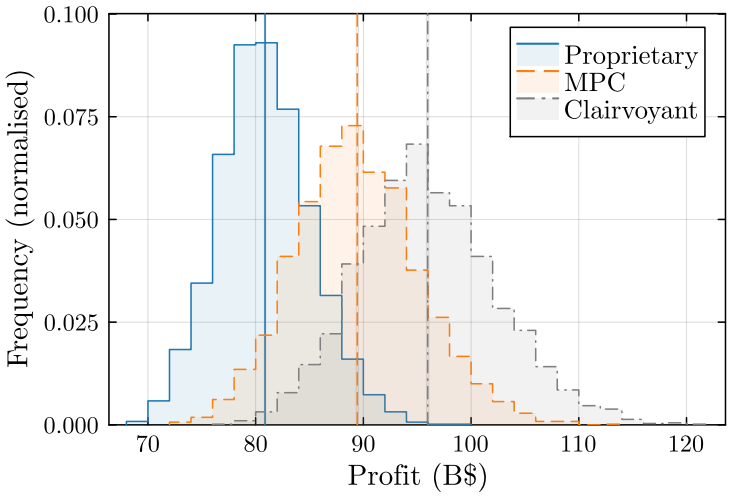

Figure 4.10: **Realised Profit.** Vertical bands present means and standard-error ranges.

Figure 4.10 shows that, on average, the MPC policy results in $\approx 11\,\%$ ($\$\,8.56$ billion USD) more profit than the proprietary policy. The benefit over the proprietary policy here is in similar proportion to that reported in [111] when comparing to a stochastic dual dynamic programming policy. Note however that this algorithm is significantly more computationally expensive than our basic MPC approach, and for negligible additional performance benefit over the proprietary policy. Furthermore, Figure 4.10 shows that, on average, the clairvoyant policy results in only $\approx 7\,\%$ more profit than the MPC policy. Since the performance of the clairvoyant policy provides an upper bound on that which is possible to attain in the real world, this shows that the MPC policy is near optimal.

For a given initial inventory and price, at the first decision-making period each policy defines a predicted future profit via the optimal value of the value-function recursion. Recall that at the beginning of a dairy season this is used to construct a forecast of the final payout milk price which is communicated to farmers ahead of time. Depending on the actual operating profit realised over the dairy season, this prediction could result in payout disappointment. Figure 4.11 graphs the simulated profit prediction–disappointment distributions of the policies.

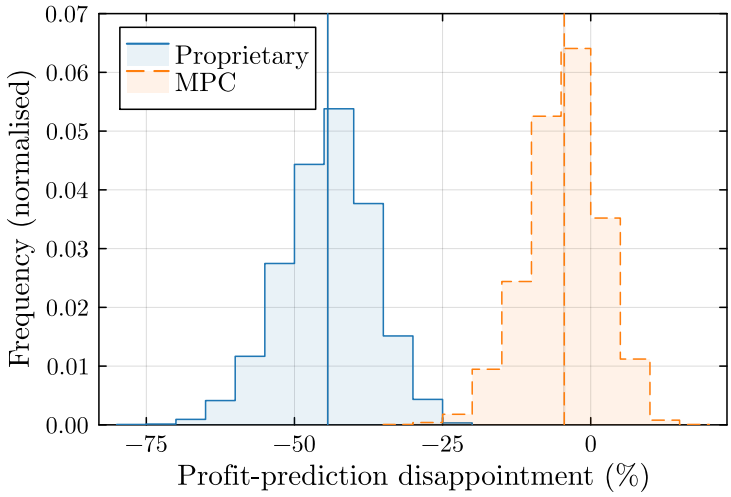

Figure 4.11: **Realised Profit-Prediction Disappointment.** Vertical bands present means and standard-error ranges.

Figure 4.11 shows that, on average, the MPC policy makes profit predictions which are $\approx 40\,\%$ more accurate than those of the proprietary policy. Although the proprietary policy results in only negative prediction disappointments (i.e., realised profit is always *more* than was predicted) and the MPC policy sometimes results in positive prediction disappointments (i.e., realised profit is *less* than was predicted), the average MPC profit-prediction disappointment is negative. This recovers the performance guarantee of Theorem 3.2. Moreover, the magnitudes of the MPC disappointments are significantly less than those of the proprietary policy, and thus the MPC policy makes more accurate profit predictions while also being conservative on average.

4.4 Discussion

All in all we conclude that MPC realises both more profits and significantly more accurate profit predictions than the proprietary policy. With MPC displaying strong out-of-sample performance using only a few years of historical data, it is well suited to the New Zealand dairy supply chain problem considered in this chapter. The inclusion of futures contracts with fixed shortage costs would be unlikely to alter this conclusion, as once offered, such contracts are deterministic. Even so, as long as milk supply and market prices are uncorrelated, it would be straightforward to incorporate uncertain milk supply via stochastic dual dynamic programming, while keeping the MPC approach for dealing with uncertain prices. Considering Examples 3.1 and 3.2 from the previous chapter, such an approach would combine the best of both worlds in terms of out-of-sample performance.

A limitation of the out-of-sample testing in this chapter is that the simulations use one stationary model for the price process. In Section 4.1 we fit a multiplicative lag-one vector-autoregressive process to the data in Figure 4.3, and this is used as the stochastic process we draw samples from to construct and then test MPC policies. However, the dynamics of the data in Figure 4.3 show variation over time. This should be expected, as a number of regulatory and demand changes occurred at different points in time throughout the fourteen-year period the data were collected; see [36, Section 9.1] and the references therein. In this way a *non*stationary model for the price process is likely to be appropriate. A data-driven decision-making policy that suitably accounts for such nonstationarity would then exhibit value over the previous MPC approach, which naïvely views historical data as stationary.

Part II

Nonstationarity

In Part I of this thesis we studied the performance of data-driven approximations of stochastic optimal control problems. Assuming that data provide the outcomes of independent and identically distributed random samples from the true distribution, we established general conditions under which sample-based approximations asymptotically converge to optimality. We also identified problem characteristics affecting the out-of-sample performance of different approximations, and used the resulting insights to develop an approximation scheme suited to a supply chain–management problem in the New Zealand dairy industry.

However, with stochastic control problems unfolding over time, in many applications it is unlikely that the underlying stochasticity responsible for generating historical data is stationary. Instead, this stochasticity may evolve over time and be *nonstationary*. In such cases, data no longer provide outcomes of random samples from the true (i.e., the current) distribution. Indeed, the dairy commodity price data in Figure 4.3 exhibits a regime change beginning around 2016. What are we to do with data from before this point in time? While older data may be less relevant to decision making in the present, they still provide *some* information that we should hope to utilise in our approximations.

Central to data-driven approximation is the estimation of probability distributions from samples. Nonstationarity requires us to rethink how we estimate such distributions. Towards our goal of understanding what drives the performance of different data-driven approximations, in this part of the thesis we study the estimation of probability distributions from nonstationary data. More specifically, we aim to estimate sets of distributions that contain the current distribution with high probability, or distributions that have a high likelihood under the historical data.

Chapter 5

Distributionally Robust Optimization with Nonstationary Data

In this chapter we study the estimation of probability distributions from nonstationary data. We treat historical observations as being generated from a time-evolving distribution whose consecutive shifts are bounded in Wasserstein distance. The standard Wasserstein distributionally robust optimization (or DRO) approach to mitigating the effects of overfitting to data would estimate a Wasserstein ambiguity ball centred at the *equally weighted* empirical distribution on the historical samples [46]. However, this is motivated by performance guarantees derived from concentration bounds for independent and identically distributed samples, which do not apply in the nonstationary case.

We address nonstationarity using a DRO model with an ambiguity set that is a Wasserstein ball centred at a *weighted* empirical distribution, thereby allowing for the time decay of past data in a way which accounts for the drift of the data-generating distribution. As for mitigating the effects of overfitting, in addition to the *bias–variance* trade-off inherent in the stationary case, our model reveals a further *variance–drift* trade-off between the numerousness of the weighted data and our exposure to distributional drift. By choosing the weights of the empirical distribution to balance this trade-off, we can estimate an ambiguity ball of distributions that contains the true (nonstationary) distribution with a high, controllable probability.

This chapter is based on the joint work:

Don't Look Back in Anger: Wasserstein
Distributionally Robust Optimization with Nonstationary Data
Dominic S. T. Keehan, Edward J. Anderson, & Wolfram Wiesemann [69].

Our work intersects with, and contributes to, two strands of the literature: DRO under nonstationary data-generating processes and DRO with heterogeneous data sources.

The literature on DRO with nonstationary data relaxes the ubiquitous identically distributed assumption by modelling temporal dependence or drift. For example, [34, 63, 114] study Wasserstein DRO models where data are generated by vector autoregressive processes with unknown parameters. Likewise, [81, 114] analyse Wasserstein DRO models where data arise from unknown finite-state Markov chains. In addition, [93] models nonstationarity via a regime-switching Wasserstein ambiguity set, and [40] investigates ϕ-divergence DRO where data are generated by certain classes of Markov chains. A common feature of this literature is the imposition of parametric or Markovian structure on the temporal dynamics. Instead, we only assume that successive data-generating distributions differ in Wasserstein distance by no more than a fixed amount, without committing to a specific parametric time-series model.

In contrast, the literature on DRO with heterogeneous data sources leverages samples from multiple distributions (sources) that are close to the data-generating (target) distribution in Wasserstein distance to construct an ambiguity set for the target. Focusing on two sources and least squares estimation, [116] proposes either interpolating between the ambiguity sets centred at each source or intersecting them. For logistic regression with two sources that are subjected to adversarial attacks, [102] intersects the two ambiguity sets. In the first work that studies more than two sources, [99] intersects Wasserstein balls centred at all sources and optimizes against the worst distribution in the intersection. The paper shows that the problem is NP-hard in general but can be solved in polynomial time when either the number of sources or the dimension of the uncertain parameters is fixed. Finally, [61] studies a dynamic setting where source reliability varies over time and proposes mechanisms to update trust in each source. The literature on DRO with heterogeneous data sources typically assumes that many samples are available per source. On the other hand, viewing each time step as a separate source, our setting provides exactly one observation per source. We will see that this renders the existing heterogeneous-source guarantees unsuited. Instead, our approach aggregates over time via weighted empirical distributions while explicitly controlling for the per-period Wasserstein drift.

This chapter proposes a Wasserstein DRO framework tailored to nonstationary data. We model the evolution of the data-generating process as a sequence of unknown distributions whose successive changes are bounded in Wasserstein distance. To derive robust decisions, we centre our ambiguity sets at *weighted* empirical distributions that downweight older samples. We then consider the fundamental variance–drift trade-off that governs data-driven decision making under nonstationarity. More specifically, our contributions are as follows.

(i) **Modelling nonstationarity via Wasserstein shifts.** We propose a DRO model that permits changes to the data-generating distribution over time, with consecutive shifts bounded in Wasserstein distance. Our ambiguity sets are centred at weighted empirical distributions, enabling the principled time decay of historical information within DRO.

(ii) **Weighted concentration for Wasserstein distances.** We extend the classical concentration bounds of [51] from equally weighted to weighted empirical distributions and obtain finite-sample guarantees for nonstationary data, with rates that explicitly reflect both the effective sample size and the magnitude of distributional drift.

(iii) **Optimizing the variance–drift trade-off.** We show how to choose weights that optimally balance statistical variance against distributional drift in our concentration bound. Our analysis covers natural weight families (e.g., time windowing and exponential decay) and characterises optimal weights within our framework.

Taken together, these contributions yield a conceptually simple and computationally practical method for decision making with nonstationary data: centre a Wasserstein ambiguity set at a weighted empirical distribution, set the weights — guided by our weighted concentration bounds — to balance variance and drift, and calibrate the ambiguity radius via cross-validation.

The remainder of this chapter is organised as follows. Section 5.1 introduces our framework. Section 5.2 develops new concentration bounds for weighted empirical distributions. Section 5.3 characterises weight choices that optimally balance variance and drift and derives optimal tuning rules for the classical schemes of time windowing and exponential smoothing. Section 5.4 reports numerical results, and Section 5.5 concludes with a discussion.

5.1 Problem Setting

We study data-driven decision problems where a decision maker observes T historical samples $\boldsymbol{\xi}_1 \sim \mathbb{P}_1, \ldots, \boldsymbol{\xi}_T \sim \mathbb{P}_T$ of an uncertain parameter $\xi \in \Xi$ governed by unknown time-varying data-generating distributions $\mathbb{P}_1, \ldots, \mathbb{P}_T \in \mathfrak{P}(\Xi)$ supported on the (known) closed set $\Xi \subseteq \mathbb{R}^m$. At time $T+1$, the decision maker selects a measurable, extended real–valued loss function $\ell : \Xi \to \mathbb{R} \cup \{\infty\}$ from an admissible class $\mathcal{L}$ to solve

$$\underset{\ell \in \mathcal{L}}{\text{minimize}} \quad \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\big[\ell(\boldsymbol{\xi})\big], \tag{5.1}$$

where $\mathcal{P} \subseteq \mathfrak{P}(\Xi)$ is the ambiguity set for the (unknown) next-period distribution $\mathbb{P}_{T+1}$. In the special case where losses are generated by a (known) cost function $f : \mathcal{X} \times \Xi \to \mathbb{R} \cup \{\infty\}$ depending on both a decision variable $x \in \mathcal{X}$ and an uncertain variable $\xi \in \Xi$, the admissible class is $\mathcal{L} = \{\xi \mapsto f(x, \xi) : x \in \mathcal{X}\}$.

When $\mathbb{P}_1 = \cdots = \mathbb{P}_{T+1}$ and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$ are drawn independently (the *stationary* setting), a common choice for the ambiguity set in (5.1) is $\mathcal{P} = \mathbb{B}_p\big(\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon\big)$; the p-Wasserstein ball of radius ε centred around the empirical distribution $\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}$. Recall that here

$$\mathbb{B}_p(\mathbb{P}; \varepsilon) \coloneqq \Big\{\mathbb{Q} \in \mathfrak{P}(\Xi) : W_p(\mathbb{P}, \mathbb{Q}) \leq \varepsilon\Big\}, \quad W_p(\mathbb{P}, \mathbb{Q}) \coloneqq \Big(\inf_{\gamma \in \Gamma(\mathbb{P}, \mathbb{Q})} \mathbb{E}_\gamma\big[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p\big]\Big)^{1/p},$$

and $\Gamma(\mathbb{P}, \mathbb{Q})$ is the set of couplings of $\mathbb{P}$ and $\mathbb{Q}$. (Throughout this chapter we will work only with the standard Euclidean norm, which we denote simply by $\|\cdot\|$.) This choice is popular because it yields tractable reformulations of the problem (5.1) for broad classes of loss functions and, under mild conditions, satisfies a finite-sample guarantee of the form

$$\Pr\Big[\mathbb{P}_{T+1} \notin \mathbb{B}_p\Big(\tfrac{1}{T}\textstyle\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon\Big)\Big] \leq c_1 \cdot \exp\Big(-c_2 T \varepsilon^m\Big),$$

for positive constants c_1 and c_2 that depend only on Ξ and p. Thus, the unknown true distribution is contained in $\mathbb{B}_p\big(\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon\big)$ with high, controllable, probability.

We depart from the stationary setting by allowing $\mathbb{P}_t$ to vary over time, but we assume that consecutive changes are bounded in Wasserstein distance. In other words, we stipulate that

$W_\infty(\mathbb{P}_t, \mathbb{P}_{t+1}) \leq \rho$ for all $t \in [T]$ and some non-negative drift bound ρ, where

$$W_\infty(\mathbb{P}, \mathbb{Q}) \coloneqq \lim_{p\to\infty} W_p(\mathbb{P}, \mathbb{Q}) = \inf_{\gamma\in\Gamma(\mathbb{P},\mathbb{Q})} \operatorname*{esssup}_{(\boldsymbol{\xi},\boldsymbol{\zeta})\sim\gamma} \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|.$$

To focus on distributional shift rather than temporal dependence, we continue to assume that, given the distributions $\mathbb{P}_1, \ldots, \mathbb{P}_T$, the samples $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$ are independent. Throughout the chapter, we assume that the drift bound ρ is known. This assumption allows our concentration bounds to explicitly capture how ρ influences the resulting probability bounds, the optimal weight choices, and the windowing and exponential smoothing schemes. At the same time, this assumption is not restrictive in practice: our results culminate in a family of optimal weighting schemes parametrised by ρ, and in applications one selects the value of ρ that best fits the data via cross-validation. We bound the distributional drift via W_∞ because it allows us to control the worst-case pointwise displacement between consecutive distributions, and it implies that $W_p(\mathbb{P}_t, \mathbb{P}_{t+1}) \leq \rho$ for all $p \in [1, \infty]$. Distributional drift bounds expressed in terms of W_p for $p < \infty$ would be interesting, but may be more challenging to work with.

In the presence of a positive drift ρ, the Wasserstein ambiguity set $\mathbb{B}_p(\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon)$ centred around the empirical distribution $\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}$ overweights earlier observations that are less representative of the next-period distribution $\mathbb{P}_{T+1}$. A seemingly natural alternative, inspired by the literature on DRO with heterogeneous sources, is to take the intersection of balls around each single-sample "data source" estimate $\mathbb{1}_{\boldsymbol{\xi}_t}$, that is, $\mathcal{P} = \bigcap_{t=1}^T \mathbb{B}_p(\mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon_t)$ for some radii $\varepsilon_1, \ldots, \varepsilon_T$, where ε_t is chosen to reflect that $\mathbb{P}_{T+1}$ is at a greater distance from earlier distributions. We next show that this construction may exhibit poor large-sample behaviour.

Example 5.1. For $\beta \in [0, 1)$, let $\varepsilon(\beta)$ denote the $(1-\beta)$-confidence p-Wasserstein radius (in the stationary setting) around an empirical distribution based on a single sample (as may be obtained from [51, Theorem 2]). If the confidence radius is set from a concentration bound, then $\varepsilon(\,\cdot\,)$ is nonincreasing and strictly positive (see also, e.g., [50]). In our time-varying setting with drift bound ρ, choose $\beta_1, \ldots, \beta_T \in [0, 1)$ such that $\sum_{t=1}^T \beta_t = \beta \in [0, 1)$ and set

$$\varepsilon_t \coloneqq \varepsilon(\beta_t) + (T - t + 1)\rho \quad \text{for all } t \in [T],$$

which achieves the guarantee [99, Proposition 4]

$$\Pr\left[\mathbb{P}_{T+1} \notin \bigcap_{t=1}^{T} \mathbb{B}_p(\mathbb{1}_{\boldsymbol{\xi}_t}; \varepsilon_t)\right] \leq \sum_{t=1}^{T} \beta_t = \beta. \tag{5.2}$$

We claim that this confidence guarantee is unsatisfactory when the drift bound is small and the sample size is large. In particular, consider the case where the data-generating process is actually stationary with $\mathbb{P}_1 = \cdots = \mathbb{P}_{T+1} = \mathbb{1}_\xi$ for some $\xi \in \Xi$. Then each $\boldsymbol{\xi}_t = \xi$, so all of the balls in (5.2) are centred around $\mathbb{1}_\xi$ and satisfy

$$\bigcap_{t=1}^{T} \mathbb{B}_p(\mathbb{1}_\xi; \varepsilon_t) = \mathbb{B}_p\Big(\mathbb{1}_\xi; \min\nolimits_{t\in[T]}\{\varepsilon_t\}\Big).$$

Since $\beta_t \leq \sum_{s=1}^{T} \beta_s = \beta$, the monotonicity and strict positivity of $\varepsilon(\,\cdot\,)$ imply that

$$\min_{t\in[T]}\{\varepsilon_t\} = \min_{t\in[T]}\big\{\varepsilon(\beta_t) + (T-t+1)\rho\big\} \geq \varepsilon(\beta) > 0,$$

and thus

$$\bigcap_{t=1}^{T} \mathbb{B}_p(\mathbb{1}_\xi; \varepsilon_t) \supseteq \mathbb{B}_p\big(\mathbb{1}_\xi; \varepsilon(\beta)\big) \neq \{\mathbb{1}_\xi\}.$$

Hence, with $\rho = 0$, even as $T \to \infty$ the intersection ambiguity set cannot collapse to $\mathbb{1}_\xi$. This is because each constituent ball is centred at a single-point empirical distribution, so the confidence radii do not decrease with T. □

The failure of the intersection-based ambiguity set to contract with T — even in the absence of drift — motivates us to centre ambiguity sets at *weighted* empirical distributions $\sum_{t=1}^{T} w_t \mathbb{1}_{\boldsymbol{\xi}_t}$. The next section develops finite-sample concentration bounds for such weighted empirical distributions under bounded drift, and Section 5.3 leverages these results to optimize the weights.

We now compare the aforementioned intersection approach to the weighted approach we will develop in the following sections. Consider a one-dimensional example with $p = 2$ and observations $\xi_1 = 1$, $\xi_2 = -1$, $\xi_3 = 2$, and $\xi_4 = 3$. Figure 5.1 shows which uniform distributions lie in the ambiguity sets of each approach. (The figure reports the means and standard deviations.) The weighted approach uses weights as per Theorem 5.2 below, and we increase the radii of the intersection approach so that the ambiguity sets are of similar size.

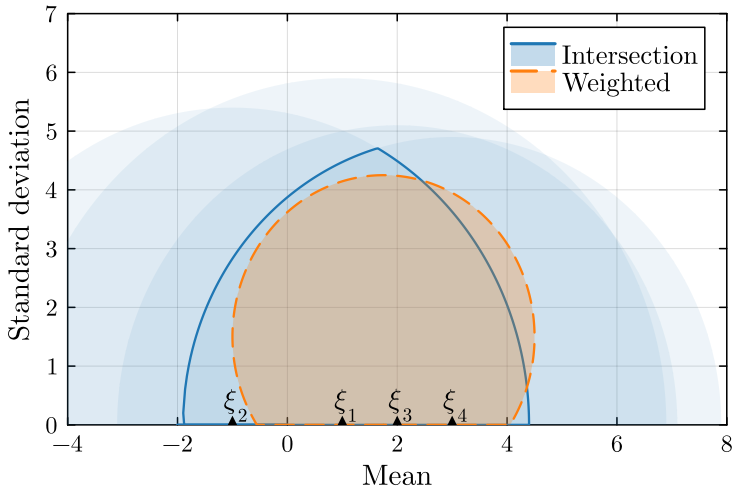

Figure 5.1: **Ambiguity Sets.** Means and standard deviations of uniform distributions under the weighted and intersection approaches for a one-dimensional example with $p = 2$.

In this example the intersection-based set is visibly more asymmetric. The weight-based set has a symmetric shape, and it can be shown that this is in fact a circle, with a vertical axis at the weighted average of the observations.

5.2 Concentration of Measure for Weighted Empiricals

The by-now classical results [51, Theorems 1 and 2] show that when $\Xi \subseteq \mathbb{R}^m$ is bounded and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N \sim \mathbb{P} \in \mathfrak{P}(\Xi)$ are independent and identically distributed random samples, then there exist constants $c_0, c_1, c_2 > 0$ and $q \in (0, 1/2]$ (depending only on m, $\operatorname{diam}(\Xi)$, and p) such that

$$\mathbb{E}\left[W_p^p\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\right)\right] \leq c_0 N^{-q} \quad \text{for all } N \in \mathbb{N}, \tag{5.3}$$

and

$$\Pr\left[W_p\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\right) \geq \varepsilon\right] \leq c_1 \cdot \exp\left(-c_2 N \varepsilon^{p/q}\right) \quad \text{for all } N \in \mathbb{N},\ \varepsilon \in \mathbb{R}_+. \tag{5.4}$$

Here $q \coloneqq \min\{p/m, 1/2\}$ when $p \neq m/2$. The boundary case $p = m/2$ entails a logarithmic correction; nevertheless, one can account for this while achieving (5.3) by choosing a q value slightly lower than $1/2$, which corresponds to a weaker bound in (5.3). Henceforth, for a given m and p, we take $q = \min\{p/m, 1/2\} - \delta$ for some small $\delta \geq 0$ so that $q \in (0, 1/2)$ achieves (5.3).

The goal of this section is to derive an analogue of (5.4) when the samples $\boldsymbol{\xi}_1 \sim \mathbb{P}_1, \ldots, \boldsymbol{\xi}_T \sim \mathbb{P}_T$ are independent but not identically distributed and when the equally weighted empirical distribution $\frac{1}{T}\sum_{t=1}^T \mathbb{1}_{\boldsymbol{\xi}_t}$ is replaced by a weighted one, $\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}$, for $w \coloneqq (w_1, \ldots, w_T) \in \mathcal{W}_T$,

where $\mathcal{W}_T$ denotes the probability T-simplex; that is, the set of non-negative T-tuples summing to 1. Our bound will explicitly reflect the drift between successive distributions and it will represent the weighting in the empirical distribution through the *effective sample size*

$$N_{\text{eff}}(w_1, \ldots, w_N) \coloneqq \frac{1}{\sum_{i=1}^{N} w_i^2}.$$

This approximates the number of equally weighted samples that would yield the same variance as the weighted sample [45, 77]. It ranges from 1 to N, with a minimal value of $N_{\text{eff}}(w) = 1$ when all of the weight is placed on a single observation and a maximal value of $N_{\text{eff}}(w) = N$ for uniform weights $w_i = 1/N$. Throughout this section, our indices and weights match the setting: we use $(i, N, w_i, \mathcal{W}_N)$ for stationary results and $(t, T, w_t, \mathcal{W}_T)$ for nonstationary results, respectively.

We derive our result in two stages. In the first stage we analyse the stationary case where the samples are drawn independently from the same distribution. Here the concentration bounds of Fournier & Guillin [51] are proved for samples with equal weights. We begin from their expectation bound and in Lemma 5.1 extend this to arbitrarily weighted empirical distributions. Then Proposition 5.1 upgrades it to a high-probability concentration bound via a McDiarmid argument. In the second stage we turn to the nonstationary case where the samples are drawn from differing distributions. Lemma 5.2 shows how the Wasserstein distance between the weighted empirical distribution formed from these nonstationary samples and the next-period distribution $\mathbb{P}_{T+1}$ relates to the Wasserstein distance between a counterpart weighted empirical distribution formed from stationary samples and $\mathbb{P}_{T+1}$. Finally, Theorem 5.1 combines Proposition 5.1 with Lemma 5.2 to deliver our main concentration bound for nonstationary data.

Lemma 5.1. *For a distribution $\mathbb{P} \in \mathfrak{P}(\Xi)$ with bounded moments, let $q \in (0, 1/2)$ achieve (5.3). Then, there exists a constant $c_0 > 0$ (depending only on m, $\mathbb{P}$'s moment bounds, p, and q) such that*

$$\mathbb{E}_{\mathbb{P}^N}\left[W_p^p\left(\sum_{i=1}^{N} w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\right)\right] \leq c_0 N_{\text{eff}}(w)^{-q} \quad \text{for all } N \in \mathbb{N},\ w \in \mathcal{W}_N.$$

Proof. Our proof proceeds in three steps. Step 1 decomposes the Wasserstein distance from the statement of the lemma into a sum of Wasserstein distances from unweighted empirical distributions. Step 2 applies the bound (5.3) for expected Wasserstein distances to these unweighted empirical distributions. Finally, in Step 3 we split the resulting sum of bounds into two parts, and show that each part is dominated by the expectation bound from the statement of the lemma.

Step 1. We reindex the weights so that $w_1 \geq \cdots \geq w_N$. This entails no loss of generality since reindexing does not change $N_{\text{eff}}(w)$. The weighted empirical distribution satisfies

$$\begin{aligned}\sum_{i=1}^{N} w_i \mathbb{1}_{\xi_i} &= w_N \sum_{i=1}^{N} \mathbb{1}_{\xi_i} + (w_{N-1} - w_N) \sum_{i=1}^{N-1} \mathbb{1}_{\xi_i} + \cdots + (w_1 - w_2)\mathbb{1}_{\xi_1} \\ &= N w_N \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\xi_i} + (N-1)(w_{N-1} - w_N) \frac{1}{N-1} \sum_{i=1}^{N-1} \mathbb{1}_{\xi_i} + \cdots + (w_1 - w_2)\mathbb{1}_{\xi_1},\end{aligned}$$

where the last expression constitutes a convex combination of the distributions $\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\xi_i}$, $\frac{1}{N-1}\sum_{i=1}^{N-1} \mathbb{1}_{\xi_i}$, …, $\mathbb{1}_{\xi_1}$ with coefficients Nw_N, $(N-1)(w_{N-1} - w_N)$, …, $(w_1 - w_2)$. Thus, by the convexity of $\mu \mapsto W_p^p(\mu, \mathbb{P})$ (see, e.g., [118, Theorem 4.8]), we obtain

$$\begin{aligned}W_p^p\Big(\textstyle\sum_{i=1}^{N} w_i \mathbb{1}_{\xi_i}, \mathbb{P}\Big) \leq N w_N W_p^p\Big(\tfrac{1}{N}\textstyle\sum_{i=1}^{N} \mathbb{1}_{\xi_i}, \mathbb{P}\Big) + (N-1)(w_{N-1} - w_N) W_p^p\Big(\tfrac{1}{N-1}\textstyle\sum_{i=1}^{N-1} \mathbb{1}_{\xi_i}, \mathbb{P}\Big) \\ + \cdots + (w_1 - w_2) W_p^p\Big(\mathbb{1}_{\xi_1}, \mathbb{P}\Big).\end{aligned}$$

Step 2. We use the expectation bound (5.3) of [51, Theorem 1] for each Wasserstein distance in the previous expression. Thus there is a constant $c > 0$ and $q \in (0, 1/2)$ (depending only on m, $\mathbb{P}$'s moment bounds, and p) such that

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}^N}\Big[W_p^p\Big(\textstyle\sum_{i=1}^{N} w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big)\Big] \\ &\leq N w_N c N^{-q} + (N-1)(w_{N-1} - w_N) c (N-1)^{-q} + \cdots + (w_1 - w_2) c \\ &= N^{1-q} w_N c + (N-1)^{1-q}(w_{N-1} - w_N) c + \cdots + (w_1 - w_2) c \\ &= \Big(N^{1-q} - (N-1)^{1-q}\Big) w_N c + \Big((N-1)^{1-q} - (N-2)^{1-q}\Big) w_{N-1} c + \cdots + w_1 c.\end{aligned}$$

Thus, we obtain that

$$\mathbb{E}_{\mathbb{P}^N}\Big[W_p^p\Big(\textstyle\sum_{i=1}^{N} w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big)\Big] \leq c \sum_{i=1}^{N} \Big(i^{1-q} - (i-1)^{1-q}\Big) w_i \leq c \sum_{i=1}^{N} i^{-q} w_i, \tag{5.5}$$

where the final inequality follows since $i^{1-q} - (i-1)^{1-q} \leq i^{-q}$ for all $i \in \mathbb{N}$ when $q > 0$.

Step 3. In (5.5) we decompose the final sum $\sum_{i=1}^{N} i^{-q} w_i$ into $H + R$ with

$$H = \sum_{i=1}^{\lfloor N_{\text{eff}}(w) \rfloor} i^{-q} w_i \quad \text{and} \quad R = \sum_{i=\lfloor N_{\text{eff}}(w) \rfloor + 1}^{N} i^{-q} w_i.$$

First consider the "head" term H. By Cauchy–Schwarz and the definition of $N_{\text{eff}}(w)$,

$$H \leq \Big(\sum_{i=1}^{\lfloor N_{\text{eff}}(w)\rfloor} i^{-2q}\Big)^{1/2}\Big(\sum_{i=1}^{\lfloor N_{\text{eff}}(w)\rfloor} w_i^2\Big)^{1/2} \leq \Big(\sum_{i=1}^{\lfloor N_{\text{eff}}(w)\rfloor} i^{-2q}\Big)^{1/2} N_{\text{eff}}(w)^{-1/2}.$$

Since $x \mapsto x^{-2q}$ is decreasing, we have

$$\sum_{i=1}^{\lfloor N_{\text{eff}}(w)\rfloor} i^{-2q} \leq 1 + \int_1^{\lfloor N_{\text{eff}}(w)\rfloor} x^{-2q}\,dx \leq \frac{\lfloor N_{\text{eff}}(w)\rfloor^{1-2q}}{1-2q},$$

and thus

$$H \leq \frac{1}{\sqrt{1-2q}}\lfloor N_{\text{eff}}(w)\rfloor^{1/2-q} N_{\text{eff}}(w)^{-1/2} \leq \frac{1}{\sqrt{1-2q}} N_{\text{eff}}(w)^{-q},$$

where the final inequality uses that $\lfloor N_{\text{eff}}(w)\rfloor \leq N_{\text{eff}}(w)$ and $1/2 - q > 0$.

Next we consider the "remainder" term R. Since $i \mapsto i^{-q}$ is decreasing and $\sum_{i=1}^N w_i = 1$, we have

$$R \leq \sum_{i=\lfloor N_{\text{eff}}(w)\rfloor+1}^{N} \Big(\lfloor N_{\text{eff}}(w)\rfloor + 1\Big)^{-q} w_i \leq \Big(\lfloor N_{\text{eff}}(w)\rfloor + 1\Big)^{-q} \leq N_{\text{eff}}(w)^{-q},$$

where the final inequality holds since $\lfloor N_{\text{eff}}(w)\rfloor + 1 \geq N_{\text{eff}}(w)$. Combining (5.5) with the bounds for H and R completes the proof with $c_0 = c(1 + 1/\sqrt{1-2q})$. □

Lemma 5.1 generalises [51, Theorem 1] to the setting of weighted empirical distributions. The convergence rates of both results coincide; the only difference is that their bound uses the actual sample size N, whereas ours uses the effective sample size $N_{\text{eff}}(w)$.

Under the additional assumption that Ξ is bounded, we can use the expectation bound of Lemma 5.1 to give a concentration bound. Note that the boundedness of Ξ implies that every distribution $\mathbb{P} \in \mathfrak{P}(\Xi)$ has bounded moments.

Proposition 5.1. *Assume* $\operatorname{diam}(\Xi) < \infty$. *For a distribution* $\mathbb{P} \in \mathfrak{P}(\Xi)$, *let* $q \in (0, 1/2)$ *be the exponent from Lemma 5.1. Then for independent and identically distributed random samples* $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N \sim \mathbb{P}$, *there exist constants* $c_1, c_2 > 0$ *(depending only on* m, $\operatorname{diam}(\Xi)$, p,

and q) such that

$$\Pr\Big[W_p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) \geq \varepsilon\Big] \leq \exp\Big(-c_1 N_{\mathrm{eff}}(w) \cdot \Big(\varepsilon^p - c_2 N_{\mathrm{eff}}(w)^{-q}\Big)_+^2\Big)$$

for all $N \in \mathbb{N}$, $w \in \mathcal{W}_N$, $\varepsilon \in \mathbb{R}_+$. (5.6)

In particular, whenever $\varepsilon \geq 2\big(c_2 N_{\mathrm{eff}}(w)^{-q}\big)^{1/p}$, we have

$$\Pr\Big[W_p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) \geq \varepsilon\Big] \leq \exp\Big(-\frac{c_1}{4} N_{\mathrm{eff}}(w) \varepsilon^{2p}\Big). \tag{5.6$'$}$$

Proof. We prove the result by applying McDiarmid's inequality to the random variable $W_p^p(\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P})$. To this end, it suffices to show that $(\xi_1, \dots, \xi_N) \mapsto W_p^p(\sum_{i=1}^N w_i \mathbb{1}_{\xi_i}, \mathbb{P})$ satisfies the bounded-differences condition, that is, to bound

$$\Big|W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\xi_i}, \mathbb{P}\Big) - W_p^p\Big(\sum_{i=1}^N w_i \mathbb{1}_{\xi_i'}, \mathbb{P}\Big)\Big|, \tag{5.7}$$

for all $(\xi_1, \dots, \xi_N), (\xi_1', \dots, \xi_N') \in \Xi^N$ with $\xi_i = \xi_i'$ for all $i \neq j$ and some $j \in [N]$.

Set $\mu' \coloneqq \sum_{i=1}^N w_i \mathbb{1}_{\xi_i'}$, and let γ' be an optimal coupling of μ' and $\mathbb{P}$, thus achieving cost $W_p^p(\sum_{i=1}^N w_i \mathbb{1}_{\xi_i'}, \mathbb{P})$. Since μ' is discrete, there exist conditional laws $\beta_i' \in \mathfrak{P}(\Xi)$ such that

$$\gamma'(\{\xi_i'\} \times \mathcal{B}) = w_i \beta_i'(\mathcal{B}) \quad \text{for all measurable } \mathcal{B} \subseteq \Xi.$$

Define a coupling γ of $\mu \coloneqq \sum_{i=1}^N w_i \mathbb{1}_{\xi_i}$ and $\mathbb{P}$ by

$$\gamma(\{\xi_i\} \times \mathcal{B}) \coloneqq w_i \beta_i'(\mathcal{B}) \quad \text{for all measurable } \mathcal{B} \subseteq \Xi \text{ and all } i \in [N].$$

Since $\xi_i = \xi_i'$ for all $i \neq j$, these slices coincide with those of γ'; only the j-th slice is relocated from $\{\xi_j'\} \times \mathcal{B}$ to $\{\xi_j\} \times \mathcal{B}$ while keeping the same conditional β_j'. With this disintegration, the cost of the optimal coupling γ' is

$$W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\xi_i'}, \mathbb{P}\Big) = \displaystyle\sum_{i=1}^N w_i \mathbb{E}_{\beta_i'}\big[\|\xi_i' - \boldsymbol{\zeta}\|^p\big].$$

For the feasible coupling γ, we analogously obtain

$$W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\xi_i}, \mathbb{P}\Big) \le \displaystyle\sum_{i\neq j} w_i \mathbb{E}_{\beta_i'}\big[\|\xi_i' - \boldsymbol{\zeta}\|^p\big] + w_j \mathbb{E}_{\beta_j'}\big[\|\xi_j - \boldsymbol{\zeta}\|^p\big].$$

Subtracting the two expressions yields

$$\begin{aligned} W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\xi_i}, \mathbb{P}\Big) - W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\xi_i'}, \mathbb{P}\Big) &\le w_j\Big(\mathbb{E}_{\beta_j'}\big[\|\xi_j - \boldsymbol{\zeta}\|^p\big] - \mathbb{E}_{\beta_j'}\big[\|\xi_j' - \boldsymbol{\zeta}\|^p\big]\Big) \\ &\le w_j \operatorname{diam}(\Xi)^p, \end{aligned}$$

since $0 \le \|\xi - \zeta\|^p \le \operatorname{diam}(\Xi)^p$ for all $\xi, \zeta \in \Xi$. A symmetric argument then bounds (5.7).

Applying McDiarmid's inequality with the bounded-differences constants $a_i \coloneqq w_i \operatorname{diam}(\Xi)^p$, so that $\sum_{i=1}^N a_i^2 = \operatorname{diam}(\Xi)^{2p} \sum_{i=1}^N w_i^2$, we obtain

$$\Pr\Big[W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) - \mathbb{E}_{\mathbb{P}^N}\Big[W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big)\Big] \ge \varepsilon\Big] \le \exp\Big(-c_1 N_{\mathrm{eff}}(w)\varepsilon^2\Big)$$
$$\text{for all } N \in \mathbb{N},\ w \in \mathcal{W}_N,\ \varepsilon \in \mathbb{R}_+,$$

where $c_1 = 2\operatorname{diam}(\Xi)^{-2p}$. Using the expectation bound of Lemma 5.1, we obtain

$$\begin{aligned} \Pr\Big[W_p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) \ge \varepsilon\Big] &= \Pr\Big[W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) \ge \varepsilon^p\Big] \\ &\le \Pr\Big[W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big) - \mathbb{E}_{\mathbb{P}^N}\Big[W_p^p\Big(\textstyle\sum_{i=1}^N w_i \mathbb{1}_{\boldsymbol{\xi}_i}, \mathbb{P}\Big)\Big] \ge \varepsilon^p - c_2 N_{\mathrm{eff}}(w)^{-q}\Big] \\ &\le \exp\Big(-c_1 N_{\mathrm{eff}}(w)\Big(\varepsilon^p - c_2 N_{\mathrm{eff}}(w)^{-q}\Big)_+^2\Big) \quad \text{for all } N \in \mathbb{N},\ w \in \mathcal{W}_N,\ \varepsilon \in \mathbb{R}_+, \end{aligned}$$

which is (5.6). Finally, if $\varepsilon^p \ge 2c_2 N_{\mathrm{eff}}(w)^{-q}$, then $\big(\varepsilon^p - c_2 N_{\mathrm{eff}}(w)^{-q}\big)_+ \ge \varepsilon^p/2$, yielding (5.6′). □

Proposition 5.1 shows that the weighted empirical distribution concentrates around its mean at an exponential rate governed by the effective sample size $N_{\mathrm{eff}}(w)$. The simplified bound (5.6′) highlights that once ε exceeds a constant multiple of $N_{\mathrm{eff}}(w)^{-q/p}$, the tail probability decays at least as fast as $\exp(-\frac{c_1}{4} N_{\mathrm{eff}}(w)\varepsilon^{2p})$.

For $p > m/2$, the bound of Proposition 5.1 matches that of [51, Theorem 2], up to replacing the actual sample size N by the effective sample size $N_{\mathrm{eff}}(w)$. By contrast, for $p < m/2$ their result yields tighter tails of order $\exp(-c_1 N \varepsilon^m)$, whereas our bound has the form $\exp(-c_1 N_{\mathrm{eff}}(w)\varepsilon^{2p})$ for larger values of ε. This discrepancy appears unavoidable when relying on McDiarmid's

inequality to move from equal to arbitrary weights.

We now turn to the nonstationary case, where the samples $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$ are drawn from differing distributions $\mathbb{P}_1, \ldots, \mathbb{P}_T$. As discussed in Section 5.1, we assume that changes between consecutive distributions are bounded in Wasserstein distance. We will make use of the following definition of *normalised weighted cumulative drift*,

$$D_p(w) \coloneqq \Big(\sum_{t=1}^{T} w_t (T - t + 1)^p \Big)^{1/p},$$

which is the weighted L_p norm of the look-back index $T - t + 1$. It ranges from 1 to T, with a minimal value of $D_p(w) = 1$ if all of the weight is placed on the most recent observation and a maximal value of $D_p(w) = T$ if all of the weight is placed on the oldest observation.

Lemma 5.2. *Let $\mathbb{P}_1, \ldots, \mathbb{P}_T, \mathbb{P}_{T+1} \in \mathfrak{P}(\Xi)$ satisfy $W_\infty(\mathbb{P}_t, \mathbb{P}_{t+1}) \leq \rho$ for all $t \in [T]$. Then for any weights $w \in \mathcal{W}_T$ and independent random samples $\boldsymbol{\xi}_t \sim \mathbb{P}_t$, $\boldsymbol{\zeta}_t \sim \mathbb{P}_{T+1}$, $t \in [T]$, there exists a coupling of the product distributions $\prod_{t=1}^T \mathbb{P}_t$ and $\prod_{t=1}^T \mathbb{P}_{T+1}$ under which*

$$W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \leq W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\zeta}_t}, \mathbb{P}_{T+1}\Big) + D_p(w)\rho \quad \textit{almost surely.}$$

Proof. By the triangle inequality for Wasserstein distances (see, e.g., [118, page 94]), we have

$$\begin{aligned} W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\xi_t}, \mathbb{P}_{T+1}\Big) &\leq W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\zeta_t}, \mathbb{P}_{T+1}\Big) + W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\xi_t}, \sum_{t=1}^T w_t \mathbb{1}_{\zeta_t}\Big) \\ &\leq W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\zeta_t}, \mathbb{P}_{T+1}\Big) + \Big(\textstyle\sum_{t=1}^T w_t \|\xi_t - \zeta_t\|^p\Big)^{1/p}. \end{aligned} \tag{5.8}$$

The triangle inequality also implies that

$$W_\infty(\mathbb{P}_t, \mathbb{P}_{T+1}) \leq W_\infty(\mathbb{P}_t, \mathbb{P}_{t+1}) + \cdots + W_\infty(\mathbb{P}_T, \mathbb{P}_{T+1}) \leq (T - t + 1)\rho \quad \text{for all } t \in [T].$$

Since the infimum in the definition of W_∞ is attained (see [55, Proposition 1]), this implies for each $t \in [T]$ that there exists a coupling $\gamma_t \in \Gamma(\mathbb{P}_t, \mathbb{P}_{T+1})$ with

$$\|\boldsymbol{\xi}_t - \boldsymbol{\zeta}_t\| \leq (T - t + 1)\rho \quad \gamma_t\text{-almost surely.}$$

The product of these couplings yields a joint coupling $\Upsilon \in \Gamma(\prod_{t=1}^T \mathbb{P}_t, \prod_{t=1}^T \mathbb{P}_{T+1})$ under which

$$\sum_{t=1}^T w_t \|\boldsymbol{\xi}_t - \boldsymbol{\zeta}_t\|^p \le \sum_{t=1}^T w_t (T-t+1)^p \rho^p = D_p^p(w)\rho^p \quad \Upsilon\text{-almost surely.}$$

Combining this with (5.8) proves the result. □

Lemma 5.2 couples the heterogeneous samples $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T$ to an independent and identically distributed proxy $\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_T$ and shows that distances to the weighted empirical distribution differ from those to the proxy by at most the addition of the weighted cumulative drift $D_p(w)\rho$. The additive term $D_p(w)\rho$ is essentially sharp: one can readily construct simple sequences of distributions for which the inequality in Lemma 5.2 is tight.

We now combine the stationary concentration bound from Proposition 5.1 with the drift bound of Lemma 5.2 to study the behaviour of weighted empirical distributions that are generated by nonstationary distributions.

Theorem 5.1. *Assume* $\operatorname{diam}(\Xi) < \infty$. *Let the distributions* $\mathbb{P}_1, \dots, \mathbb{P}_T, \mathbb{P}_{T+1} \in \mathfrak{P}(\Xi)$ *satisfy* $W_\infty(\mathbb{P}_t, \mathbb{P}_{t+1}) \le \rho$ *for all* $t \in [T]$, *and let* q *be the exponent from Lemma 5.1 for* $\mathbb{P}_{T+1}$. *Then for independent random samples* $\boldsymbol{\xi}_t \sim \mathbb{P}_t$, $t \in [T]$, *there exist constants* $c_1, c_2 > 0$ *(depending only on* m, $\operatorname{diam}(\Xi)$, p, *and* q*) such that*

$$\Pr\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \ge \varepsilon\Big] \le \exp\Big(-c_1 N_{\mathrm{eff}}(w) \cdot \Big((\varepsilon - D_p(w)\rho)_+^p - c_2 N_{\mathrm{eff}}(w)^{-q}\Big)_+^2\Big)$$
$$\text{for all } T \in \mathbb{N},\ w \in \mathcal{W}_T,\ \varepsilon \in \mathbb{R}_+. \quad (5.9)$$

In particular, whenever $\varepsilon \ge D_p(w)\rho + 2(c_2 N_{\mathrm{eff}}(w)^{-q})^{1/p}$, *we have*

$$\Pr\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \ge \varepsilon\Big] \le \exp\Big(-\frac{c_1}{4} N_{\mathrm{eff}}(w) \cdot \big(\varepsilon - D_p(w)\rho\big)_+^{2p}\Big). \quad (5.9')$$

Proof. For independent random samples $\boldsymbol{\xi}_t \sim \mathbb{P}_t$, $\boldsymbol{\zeta}_t \sim \mathbb{P}_{T+1}$, $t \in [T]$, Lemma 5.2 provides a coupling Υ of the two T-product distributions under which

$$W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \le W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\zeta}_t}, \mathbb{P}_{T+1}\Big) + D_p(w)\rho \quad \text{almost surely.}$$

Hence,

$$\begin{aligned}\Pr\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \geq \varepsilon\Big] &= \Upsilon\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \geq \varepsilon\Big] \\ &\leq \Upsilon\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\zeta}_t}, \mathbb{P}_{T+1}\Big) \geq \big(\varepsilon - D_p(w)\rho\big)_+\Big] \\ &= \Pr\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\zeta}_t}, \mathbb{P}_{T+1}\Big) \geq \big(\varepsilon - D_p(w)\rho\big)_+\Big] \quad \text{for all } T \in \mathbb{N},\ w \in \mathcal{W}_T,\ \varepsilon \in \mathbb{R}_+.\end{aligned}$$

Applying Proposition 5.1 to the right-hand side yields (5.9). For the tail (5.9′), note that if $\varepsilon \geq D_p(w)\rho + 2\big(c_2 N_{\mathrm{eff}}(w)^{-q}\big)^{1/p}$, then $\big((\varepsilon - D_p(w)\rho)^p - c_2 N_{\mathrm{eff}}(w)^{-q}\big)_+ \geq \frac{1}{2}(\varepsilon - D_p(w)\rho)_+^p$. □

Theorem 5.1 shows that, after accounting for $D_p(w)\rho$ and $N_{\mathrm{eff}}(w)^{-q/p}$, the weighted empirical distribution concentrates about $\mathbb{P}_{T+1}$. When $\rho = 0$ and thus $D_p(w)\rho = 0$, the bound reduces to that of Proposition 5.1. The comparison with [51, Theorem 2] carries over mutatis mutandis from Proposition 5.1: when $p > m/2$ our tails match theirs (up to replacing N by $N_{\mathrm{eff}}(w)$), whereas for $p < m/2$ their sharper exponent m persists, while our bound has the exponent $2p$ and is effective only above $N_{\mathrm{eff}}(w)^{-q/p}$.

5.3 Optimal Observation Weightings

Theorem 5.1 shows that, for a given ambiguity radius ε, different choices of observation weights w lead to different probabilistic guarantees that the next-period distribution $\mathbb{P}_{T+1}$ lies within Wasserstein distance ε of the weighted empirical distribution $\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}$. It is therefore natural to seek a weighting that achieves the strongest probabilistic guarantee at this ambiguity radius, thereby providing the least conservative ambiguity set.

For simplicity we focus on the tail regime of Theorem 5.1 where (5.9′) applies. This corresponds to values of ε that are not too small, and the bound simplifies to

$$\Pr\Big[W_p\Big(\textstyle\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\Big) \geq \varepsilon\Big] \leq \exp\Big(-\frac{c_1}{4} N_{\mathrm{eff}}(w) \cdot \Big(\varepsilon - D_p(w)\rho\Big)_+^{2p}\Big).$$

In this regime, minimizing the conservatism of the ambiguity set is equivalent to maximizing the magnitude of the exponent on the right-hand side. Accordingly, we formulate the problem of selecting the optimal weighting w as

$$\underset{w \in \mathcal{W}_T}{\text{maximize}} \quad N_{\mathrm{eff}}(w) \cdot \Big(\varepsilon - D_p(w)\rho\Big)_+^{2p}. \tag{5.10}$$

This objective presents a fundamental *variance–drift* trade-off between two opposing forces when choosing the weights w: on the one hand, *reducing sampling variance* by increasing the effective sample size $N_{\mathrm{eff}}(w)$; and on the other, *reducing distributional drift* by decreasing the weighted cumulative drift $D_p(w)\rho$. Putting more weight on recent samples reduces our exposure to drift but limits the number of effectively independent random samples, and vice versa. Note that dividing the objective in (5.10) by ρ^{2p} implies that the optimal solutions depend solely on the ratio ε/ρ rather than on the individual values of ε and ρ.

Problem (5.10) is high-dimensional and nonconvex, and thus appears to be computationally challenging. Fortunately, it admits a significant simplification: it always possesses an optimal solution that lies within a simple two-parameter family of weights which are monotone truncated polynomials in the look-back index.

Theorem 5.2. *The problem* (5.10) *admits an optimal solution*

$$w_t = \Big(a_1 - a_2(T-t+1)^p\Big)_+, \quad t \in [T],$$

for some constants $a_1, a_2 \geq 0$.

Proof. The positive-part operation in problem (5.10) must be inactive at optimality; otherwise any feasible w would be optimal. We may thus assume $\varepsilon > D_p(w)\rho$ and equivalently maximize

$$N_{\mathrm{eff}}(w)\Big(\varepsilon - D_p(w)\rho\Big)^{2p}$$

over $w \in \mathcal{W}_T = \{w \in \mathbb{R}^T : \sum_{t=1}^T w_t = 1,\, w_t \geq 0,\, t \in [T]\}$. Since the objective function is continuous and the feasible region is compact, optimal solutions exist. Moreover, the linear independence constraint qualification holds at every feasible point: if $\mathcal{A} = \{t \in [T] : w_t = 0\}$ is the active set, then $\mathcal{A} \subsetneq [T]$ since at least one weight must be strictly positive, and thus the gradients of the active inequalities together with the equality-constraint gradient are linearly independent. Consequently, the Karush–Kuhn–Tucker (KKT) conditions are necessary for local (and hence global) optimality. We claim that any KKT point must exhibit the structure stated in the theorem.

Introducing a multiplier λ for the equality constraint and multipliers $\mu_t \geq 0$ for the inequalities in $\mathcal{W}_T$, the Lagrangian reads

$$L(w, \lambda, \mu) = N_{\mathrm{eff}}(w)\Big(\varepsilon - D_p(w)\rho\Big)^{2p} + \lambda\Big(\sum_{t=1}^T w_t - 1\Big) + \sum_{t=1}^T \mu_t w_t.$$

By the definitions of $N_{\text{eff}}(w)$ and $D_p(w)$, we have

$$\frac{\partial N_{\text{eff}}(w)}{\partial w_s} = -2w_s N_{\text{eff}}(w)^2 \quad \text{and} \quad \frac{\partial D_p(w)}{\partial w_s} = \frac{(T-s+1)^p}{p} D_p(w)^{1-p}.$$

For strictly positive weights $w_s > 0$, complementary slackness implies that $\mu_s = 0$, and hence the stationarity condition $\frac{\partial}{\partial w_s} L(w, \lambda, \mu) = 0$ reduces to

$$-2w_s N_{\text{eff}}(w)^2 \Big(\varepsilon - D_p(w)\rho\Big)^{2p} - 2\rho N_{\text{eff}}(w) \Big(\varepsilon - D_p(w)\rho\Big)^{2p-1} (T-s+1)^p D_p(w)^{1-p} + \lambda = 0.$$

Dividing through by $2N_{\text{eff}}(w)^2(\varepsilon - D_p(w)\rho)^{2p-1}$ yields $w_s = a_1 - a_2(T-s+1)^p$, where

$$a_1 = \frac{\lambda}{2N_{\text{eff}}(w)^2(\varepsilon - D_p(w)\rho)^{2p}} \quad \text{and} \quad a_2 = \frac{\rho D_p(w)^{1-p}}{N_{\text{eff}}(w)(\varepsilon - D_p(w)\rho)}. \tag{5.11}$$

For indices where w_s would be negative, the complementary slackness condition enforces $w_s = 0$. The constants $a_1, a_2 \geq 0$ are determined by the constraint $\sum_{t=1}^{T} w_t = 1$ and the threshold beyond which the weights vanish. Hence, every locally optimal weighting takes the claimed form. □

Theorem 5.2 shows that the optimal weights decay polynomially when looking backwards in time, and that they are truncated to zero beyond a finite horizon. Figure 5.2 shows how these optimal weightings behave in a specific example.

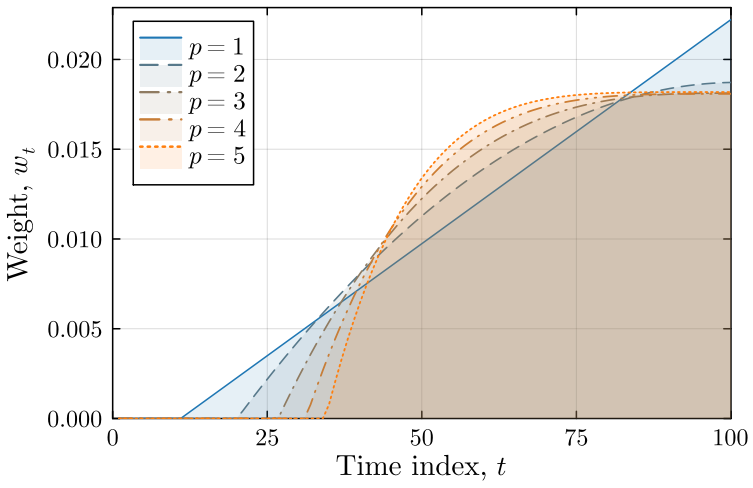

Figure 5.2: **Optimal Weightings for Different Orders p.** Here $\varepsilon/\rho = 90p$ and $T = 100$.

The fact that weight concentrates on the most recent observations as p increases can be explained by the geometry of the Wasserstein distances: the exponent p affects how severely large transportation distances are penalised. In determining the best variance–drift trade-off, larger

values of p increase the sensitivity of the W_p distance to the oldest (most drifted) samples. Optimal solutions therefore concentrate weight on a recent subset of samples to avoid heavy penalties.

Theorem 5.2 also yields an efficient numerical solution method for (5.10). Without loss of generality, we may restrict ourselves to weightings of the form in Theorem 5.2 that are supported on the $s \in [T]$ most recent observations. For each fixed s, the constraint $\sum_{t=1}^{T} w_t = 1$ fixes $a_1 = (1 + a_2 \sum_{t=1}^{s} t^p)/s$, so the feasible set on that support collapses to a one-dimensional curve parametrised by a_2. Enforcing truncation exactly at s (i.e., $w_{T-s+1} > 0$ and $w_{T-s} = 0$) yields a single admissible interval $a_2 \in \big[1/(s(s+1)^p - \sum_{t=1}^{s} t^p), 1/(s^{p+1} - \sum_{t=1}^{s} t^p)\big)$. Thus, the global search for optimal weights reduces to a finite union of one-dimensional line searches.

The special case of $p = 1$ allows for a closed-form solution.

Proposition 5.2. *Assume $\varepsilon/\rho > 1$ and let $p = 1$. Then problem (5.10) has an optimal solution supported on the $s = \lfloor \varepsilon/\rho \rfloor$ most recent observations given by*

$$w_t = \begin{cases} \dfrac{2((\varepsilon/\rho) + t - T - 1)}{s(2\varepsilon/\rho - s - 1)} & \text{if } t \geq T - s + 1, \\ 0 & \text{if } t \leq T - s. \end{cases} \tag{5.12}$$

Proof. Theorem 5.2 implies that any optimal solution to problem (5.10) with $p = 1$ has a truncated affine form, so that the weights are supported on a contiguous block of the most recent observations. Thus there exists an integer $s \in [T]$ such that

$$w_t = a_1 - a_2(T - t + 1) \quad \text{for } t \in \{T - s + 1, \ldots, T\}, \qquad w_t = 0 \text{ otherwise.}$$

Here the value of w_t at $t = T - s$ would be negative if it were not truncated. Hence $a_1 - a_2 s \geq 0$ and $a_1 - a_2(s+1) < 0$, i.e., $s = \lfloor a_1/a_2 \rfloor$. Under this choice of s,

$$D_1(w) = \sum_{t=1}^{T} w_t(T - t + 1) = \sum_{t=T-s+1}^{T} w_t \frac{a_1 - w_t}{a_2} = \frac{a_1}{a_2} - \frac{1}{a_2} \sum_{t=1}^{T} w_t^2, \tag{5.13}$$

where the second equality replaces $(T - t + 1)$ with $(a_1 - w_t)/a_2$, and the third uses $\sum_{t=1}^{T} w_t = 1$.

From (5.11) in the proof of Theorem 5.2 we know that

$$a_2 = \frac{\rho}{N_{\text{eff}}(w)(\varepsilon - D_1(w)\rho)}.$$

Plugging in $N_{\text{eff}}(w) = \left(\sum_{t=1}^{T} w_t^2\right)^{-1}$ and the expression (5.13) for $D_1(w)$, cross-multiplying yields

$$a_2\Big(\varepsilon - \frac{a_1\rho}{a_2}\Big) + \rho\sum_{t=1}^{T} w_t^2 = \rho\sum_{t=1}^{T} w_t^2.$$

We thus deduce that $a_1/a_2 = \varepsilon/\rho$, giving $s = \lfloor a_1/a_2 \rfloor = \lfloor \varepsilon/\rho \rfloor$.

Since the weights on the last s periods all satisfy $w_t = a_1 - a_2(T-t+1)$ and $\sum_{t=1}^{T} w_t = 1$, we have $sa_1 - \frac{1}{2}a_2 s(s+1) = 1$. Then, substituting $a_1 = (\varepsilon/\rho)a_2$ into this equation, we deduce that $a_2 s\big((\varepsilon/\rho) - (s+1)/2\big) = 1$, and hence

$$a_2 = \frac{2\rho}{s(2\varepsilon - \rho(s+1))}.$$

Substituting this expression and $a_1 = (\varepsilon/\rho)a_2$ into $w_t = a_1 - a_2(T-t+1)$ proves the result. □

Proposition 5.2 shows that for $p = 1$ the optimal weights increase linearly with recency. When ε/ρ is an integer, then this observation has zero weight. So it is the last $\lceil \varepsilon/\rho \rceil - 1$ periods that have positive weight.

We now consider two classical schemes for assigning observation weights. The first is *windowing*, in which equal weights are assigned to the most recent s observations and older data are discarded. A window of size $s \in [T]$ therefore sets

$$w_T = \cdots = w_{T-s+1} = \frac{1}{s}, \quad w_{T-s} = \cdots = w_1 = 0.$$

The second is *smoothing*, in which weights decay geometrically backward in time at a fixed rate. Parametrised by a decay rate $\alpha \in [0,1]$, this sets

$$w_T = \alpha,\ w_{T-1} = \alpha(1-\alpha),\ w_{T-2} = \alpha(1-\alpha)^2,\ \ldots,$$

where we truncate to the available history length and scale the sum to 1.[1]

Windowing and smoothing are simple and widely used in dealing with stochasticity, but they can also be treated in our distributionally ambiguous theoretical framework. The following result summarises the optimal parameter choices for $p = 1$ and shows their relationship with ε/ρ.

Proposition 5.3. *Let $p = 1$. Then:*

[1] For $\alpha = 0$, we take the limiting form as $\alpha \to 0$, corresponding to the equal weighting $w_T = \cdots = w_1 = 1/T$.

(i) *The optimal window size for problem* (5.10) *is either*

$$s = \left\lfloor \frac{1}{3}\Big(2\varepsilon/\rho - 1\Big) \right\rfloor_{[1,T]} \quad \text{or} \quad s = \left\lceil \frac{1}{3}\Big(2\varepsilon/\rho - 1\Big) \right\rceil_{[1,T]}.$$

(ii) *As* $T \to \infty$*, the optimal smoothing decay rate for problem* (5.10) *is*

$$\alpha = \left(\frac{3}{\varepsilon/\rho + 1}\right)_{\left[(\rho/\varepsilon)_{[0,1]},1\right]}.$$

Proof. (i): For windowing with $s \in [T]$, problem (5.10) becomes

$$\underset{s\in[T]}{\text{maximize}} \quad f(s) \coloneqq s\Big(\varepsilon - \tfrac{1}{s}\textstyle\sum_{t=1}^{s} t\rho\Big)_+^2 = s\Big(\varepsilon - \tfrac{1}{2}(s+1)\rho\Big)_+^2.$$

Here the function $f(s) = 0$ for $s \geq 2\varepsilon/\rho - 1$, and thus we equivalently

$$\underset{s\in[T]}{\text{maximize}} \quad g(s) \coloneqq s\Big(\varepsilon - \tfrac{1}{2}(s+1)\rho\Big)^2 \quad \text{subject to} \quad s \leq 2\varepsilon/\rho - 1. \tag{5.14}$$

The function g is initially increasing and has at most one turning point on $(-\infty, 2\varepsilon/\rho - 1]$. Differentiating, it has the unconstrained first-order optimality condition

$$\frac{dg(s)}{ds} = \Big(\varepsilon - \tfrac{1}{2}(s+1)\rho\Big)^2 - s\rho\Big(\varepsilon - \tfrac{1}{2}(s+1)\rho\Big) = 0,$$

which has solutions $\frac{1}{3}(2\varepsilon/\rho - 1)$ and $2\varepsilon/\rho - 1$. The former maximizes g on $(-\infty, 2\varepsilon/\rho - 1]$. Since the function g is univariate, we conclude that one of the adjacent integers within $[T]$ solves (5.14).

(ii): For smoothing with $\alpha \in (0, 1]$, as $T \to \infty$ the objective function in (5.10) becomes

$$\Big(\textstyle\sum_{t=1}^{\infty}\big(\alpha(1-\alpha)^{t-1}\big)^2\Big)^{-1}\Big(\varepsilon - \textstyle\sum_{t=1}^{\infty}\alpha(1-\alpha)^{t-1}t\rho\Big)_+^2 = \Big(\tfrac{2}{\alpha} - 1\Big)\Big(\varepsilon - \tfrac{1}{\alpha}\rho\Big)_+^2.$$

(Here we have used a reversal of indexing.) Maximizing in α, the proof is similar to that of (i). □

Proposition 5.3(i) offers a natural interpretation. Ignoring the integer projection, the optimal window size grows approximately linearly with ε/ρ: doubling the radius ε of the Wasserstein ball roughly doubles the optimal look-back horizon. Intuitively, larger ambiguity radii tolerate greater distributional drift, making it beneficial to use older (less relevant) observations.

The optimal weightings identified by Theorem 5.2 yield the smallest (least conservative)

radius ε that ensures coverage of the next-period distribution $\mathbb{P}_{T+1}$ at a prescribed confidence level $1-\beta \in (0,1]$. In the stationary case ($\rho = 0$), this radius tends to zero as T increases. In contrast, when $\rho > 0$, the ambiguity radius ε cannot vanish — its size must reflect the drift. Our next result quantifies how the minimal confidence radius scales with ρ (and β); for clarity we focus on the case $p = 1$.

Proposition 5.4. *Under the assumptions of Theorem 5.1, let $p = 1$. For any $\beta \in (0,1)$, there exist constants $c_1, c_2, c_3, c_4 > 0$ (independent of T, ρ, and β), such that, if the history length $T \geq \left(\frac{12}{c_2}\rho^{-2}\log(1/\beta)\right)^{1/3}$ and the weights $w \in \mathcal{W}_T$ are optimally chosen, then the minimal ambiguity radius $\varepsilon(\beta,\rho)$ for which*

$$\Pr\left[W_1\left(\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\right) \geq \varepsilon(\beta,\rho)\right] \leq \beta,$$

is given by

$$\varepsilon(\beta,\rho) = \begin{cases} c_1\Big(\rho \cdot \log(1/\beta)\Big)^{1/3} + c_2\Big(\rho^{2q} \cdot \log(1/\beta)^{-q}\Big)^{1/3} & \text{if } \rho < \rho^\star(\beta), \\ \rho + c_3 \cdot \log(1/\beta)^{1/2} + c_4 & \text{if } \rho \geq \rho^\star(\beta), \end{cases}$$

for a threshold $\rho^\star(\beta) = \left(\frac{12}{c_2}\log(1/\beta)\right)^{1/2}$.

Proof. With $p = 1$, Theorem 5.1 yields

$$\Pr\left[W_1\left(\sum_{t=1}^T w_t \mathbb{1}_{\boldsymbol{\xi}_t}, \mathbb{P}_{T+1}\right) \geq \varepsilon\right] \leq \exp\left(-c_2 N_{\text{eff}}(w)\Big(\big(\varepsilon - D_1(w)\rho\big)_+ - c_1 N_{\text{eff}}(w)^{-q}\Big)_+^2\right). \quad (5.15)$$

Our goal is to find the smallest value of ε for which the the right-hand side of (5.15) is at most β. We assume $T \geq \left(\frac{12}{c_2}\rho^{-2}\log(1/\beta)\right)^{1/3}$ and use triangular weights $w(s) \in \mathcal{W}_T$ supported on the most recent $s \in [T]$ observations, as given by Proposition 5.2. A direct calculation shows that

$$D_1(w(s)) = \frac{s+2}{3} \quad \text{and} \quad N_{\text{eff}}(w(s)) = \frac{3s(s+1)}{2(2s+1)} \geq \frac{3s}{4}.$$

The bound (5.15) is at most β once

$$\big(\varepsilon - D_1(w(s))\rho\big)_+ - c_1 N_{\text{eff}}(w(s))^{-q} \geq \left(\frac{\log(1/\beta)}{c_2 N_{\text{eff}}(w(s))}\right)^{1/2}, \quad (5.16)$$

since in that case the exponent on the right-hand side of (5.15) does not exceed $-\log(1/\beta)$. Note

that $D_1(w(s))$ and $N_{\mathrm{eff}}(w(s))$ both increase with s. To find the smallest ε for which (5.16) holds, we choose s to balance these terms, analysing the cases $\rho < \rho^\star(\beta)$ and $\rho \geq \rho^\star(\beta)$ separately.

Case 1: $\rho < \rho^\star(\beta)$. Set

$$s^\star = \left(\frac{12}{c_2}\rho^{-2}\log(1/\beta)\right)^{1/3} \quad \text{and} \quad s = \lceil s^\star \rceil.$$

Our earlier assumption on T ensures that $s \in [T]$. Since $D_1(w(s)) \leq (s^\star + 3)/3$, we have

$$D_1(w(s))\rho \leq \frac{s^\star}{3}\rho + \rho \leq \frac{4}{3}\left(\frac{12}{c_2}\rho\log(1/\beta)\right)^{1/3},$$

where the second inequality uses that $\rho < \rho^\star(\beta)$. Moreover,

$$\left(\frac{\log(1/\beta)}{c_2 N_{\mathrm{eff}}(w(s))}\right)^{1/2} \leq \left(\frac{4\log(1/\beta)}{3c_2 s^\star}\right)^{1/2} = \frac{2}{\sqrt{3}}\left(\frac{\log(1/\beta)\rho}{\sqrt{12}c_2}\right)^{1/3}.$$

Define

$$c_3 \coloneqq \frac{4}{3}\left(\frac{12}{c_2}\right)^{1/3} + \frac{2}{\sqrt{3}}\left(\frac{1}{\sqrt{12}c_2}\right)^{1/3} \quad \text{and} \quad c_4 \coloneqq c_1\left(\frac{4}{3}\right)^q\left(\frac{12}{c_2}\right)^{-q/3},$$

and set

$$\varepsilon(\beta,\rho) = c_3\left(\rho\log(1/\beta)\right)^{1/3} + c_4\left(\rho^{2q}\log(1/\beta)^{-q}\right)^{1/3}.$$

Then

$$\varepsilon(\beta,\rho) - D_1(w(s))\rho \geq \left(\frac{\log(1/\beta)}{c_2 N_{\mathrm{eff}}(w(s))}\right)^{1/2} \quad \text{and} \quad c_4\left(\rho^{2q}\log(1/\beta)^{-q}\right)^{1/3} \geq c_1 N_{\mathrm{eff}}(w(s))^{-q},$$

using that $N_{\mathrm{eff}}(w(s)) \geq 3s^\star/4$. We conclude that this choice of s and $\varepsilon(\beta,\rho)$ satisfies (5.16), and therefore that the right-hand side of (5.15) is bounded above by β.

Case 2: $\rho \geq \rho^\star(\beta)$. We set $s = 1$ so that $D_1(w) = 1$ and $N_{\mathrm{eff}}(w(s)) = 1$, and with

$$\varepsilon(\beta, \rho) = \rho + \left(\frac{\log(1/\beta)}{c_2}\right)^{1/2} + c_1,$$

the inequality (5.16) holds as an equality, and hence the right-hand side of (5.15) equals β.

In each case we have established the result after an appropriate relabelling of the constants. □

Since $q \in (0, 1/2)$, Proposition 5.4 implies that, for small drift, the minimal confidence radius scales as at most $\rho^{1/3}$. For large drift, the optimal weighting strategy is to place all weight on the most recent observation, the drift dominates, and the minimal confidence radius scales as ρ. For comparison, using the intersection-of-balls approach, a union-bound argument shows that the radius of the t-th ball must satisfy a lower bound of the form $(T-t+1)\rho + c$ to achieve $1-\beta$ confidence (see also [99, Proposition 4]). This lower bound scales linearly in both T and ρ, whereas our weighted scheme achieves a radius that is effectively independent of T and that scales as $\rho^{1/3}$ in the small-drift regime, becoming linear in ρ only when the drift is large.

5.4 Numerical Experiments

In this section we study a decision problem faced by a newsvendor supplying a random demand that evolves over time. We approach the problem using DRO and compare our weight-based approach to the intersection-based approach of [99]. We also compare to SAA and smoothing.

Consider a newsvendor with underage cost $c_{\mathrm{u}} > 0$ and overage cost $c_{\mathrm{o}} > 0$ supplying a random demand $\boldsymbol{\xi}$ supported on $\Xi \subseteq \mathbb{R}$. For an ambiguity set of demand distributions $\mathcal{P} \subseteq \mathfrak{P}(\Xi)$, the DRO order quantity solves

$$\underset{x \in \Xi}{\text{minimize}} \quad \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}_{\mathbb{Q}}\Big[c_{\mathrm{u}}(\boldsymbol{\xi} - x)_+ + c_{\mathrm{o}}(x - \boldsymbol{\xi})_+\Big]. \tag{5.17}$$

Here $\mathcal{P}$ may be a Wasserstein ball centred around a weighted empirical distribution, or an intersection of multiple Wasserstein balls each centred around their own empirical distributions. Since the newsvendor cost function can be expressed as the pointwise maximum of a finite number of affine functions, in the aforementioned cases the DRO problem can be reformulated as a convex-cone program following [99, Corollary 2]. In what follows we use $p = 2$ for the Wasserstein ball–based ambiguity sets.

In our experiments we set $c_{\mathrm{u}} = 4$ and $c_{\mathrm{o}} = 1$. For the demand process we use a mixture of binomial distributions and at each time t set

$$\boldsymbol{\xi}_t \sim 0.9 \cdot \mathrm{Binomial}(1000, p_t) + 0.1 \cdot \mathrm{Binomial}(1000, q_t),$$

where p_t and $q_t \in [0, 1]$ are time varying. This amounts to an economy with 1000 consumers and two possible configurations, capturing switches from surplus to shortage. Each consumer independently demands 1 unit with probability p_t or q_t, so $\boldsymbol{\xi}_t$ is supported on $\Xi = [0, 1000]$. Here p_t and q_t evolve independently over time from initial values $p_1 = 0.1$ and $q_1 = 0.5$, following random walks with Triangular$(-\delta, \delta, 0)$ innovations. (We project these values onto $[0, 1]$ to ensure they remain valid probabilities.) The value of $\delta \geq 0$ parametrises the extent of nonstationarity.

To assess performance, we use a training-and-testing approach, averaging over 1000 simulations. Each simulation uses $T = 100$ total historical demand observations. Newsvendor orders are placed from times $t = 71$ to 100 using prior observations, and costs are based on the demand in the next period. The parameter values that minimize this total training cost are then used in a final ordering decision, which incurs a testing cost. This is computed by using an analytic formula for the expected newsvendor loss under binomial demand, combined with an additional average over 1000 simulated jumps from time $t = 100$ to 101.

Our weight-based approach for the ambiguity set in (5.17) uses two parameters, ε and ρ, with the observation weights in the central empirical distribution chosen according to the characterisation of Theorem 5.2. For the intersection-based approach, which uses constituent ambiguity balls centred around each observation, we again use the two parameters ε and ρ and set the radius of the t-th ball to $\varepsilon + (T - t + 1)\rho$. (When the realisations of the observations are such that the intersection is empty, we increase the radii until it is nonempty.) SAA and smoothing consider only the equally weighted empirical and the decay rate–parametrised weighted empirical on the observations, respectively.

Table 5.1 shows the parameter ranges searched over when training each approach. Here for scalars $a \leq b$ and a positive integer n, we use $\texttt{LinRange}(a, b; n)$ for the set of n equally spaced values between a and b, inclusive. Similarly, for scalars $a \leq b \in (0, \infty)$ and a positive integer n, we use $\texttt{LogRange}(a, b; n)$ for the set of n values forming a geometric sequence between a and b, inclusive. This logarithmic spacing is the natural choice for parametrising the decay rate α and the ratio ρ/ε.

Table 5.1: **Parameter Ranges.**

Decay rate	$\alpha \in \{0\} \cup \texttt{LogRange}(10^{-4}, 1; 30)$
Ambiguity radius	$\varepsilon \in \{0\} \bigcup_{i=0}^{2} \texttt{LinRange}(10^{i}, 10^{i+1}; 10)$
Shift bound	$\rho : \rho/\varepsilon \in \{0\} \cup \texttt{LogRange}(10^{-4}, 1; 30)$[2]

We formulate problem (5.17) as a convex-cone program in `Julia` [19] using `JuMP.jl` [83] and `MathOptInterface.jl` [80], and solve it with `Gurobi` [62]. Figure 5.3 graphs the average test performance of each approach for different extents of nonstationarity.

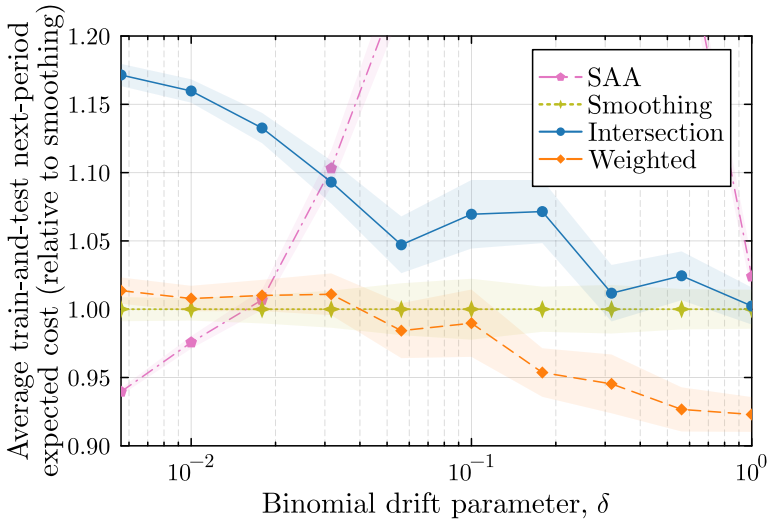

Figure 5.3: **Expected Out-of-Sample Performance.** Bands present standard-error ranges.

For a small extent of nonstationarity, unsurprisingly, SAA performs the best, while the intersection approach performs the worst as it cannot recover the unweighted empirical (see Example 5.1). Even though smoothing can recover SAA for $\alpha = 0$, it performs worse than SAA due to the effects of training. In the same way, smoothing performs slightly better than our weighted approach, as it uses only one parameter (compared to two), making it easier to train. As the extent of nonstationarity increases, SAA performs the worst, and our weighted approach performs the best. The intersection approach continues to perform worse than our weighted approach, as the radii required to achieve a prescribed confidence level depend on the history length, whereas Proposition 5.4 shows that this is not the case for our weighted approach. (Although we note that Proposition 5.4 is proved in the different context of $p = 1$.)

[2]For $\varepsilon = 0$ we treat the ratio ρ/ε separately.

5.5 Discussion

This chapter introduces a nonparametric framework for DRO with nonstationary data. By placing weighted empirical distributions at the centre of Wasserstein balls and deriving new concentration bounds, we provide a principled approach to trading off statistical variance against nonstationarity.

We view our contribution as a first step that opens up several promising directions for future research. Firstly, our concentration results bound the probability that the unknown data-generating distribution lies within the ambiguity set, and they therefore inherit the curse of dimensionality. It would be valuable to investigate whether our analysis can be extended to dimension-free concentration arguments based on the equivalences between DRO and regularization [22, 52].

Secondly, our study focuses on Wasserstein distances. There is a rich literature on alternative discrepancy measures between probability distributions, such as ϕ-divergences, which are well suited for distributions with discrete support [12, 24]. The notion of nonstationarity with bounded per-period drift applies equally to these alternative discrepancy measures, and it is natural to ask what the corresponding analogues of our results would be in these settings.

Thirdly, our analysis assumes that the historical samples are independent. An important open question is whether our results can be generalised to settings with controlled dependence, in analogy to classical generalisations of the central limit theorem [21, 43]. We expect that such extensions are possible, although technically involved.

Lastly, we compare our weighted approach to the intersection approach of [99], for which each source distribution at each time period has an ambiguity set centred at the historical sample from that time period. In principle however, it would be possible to use a hybrid approach, where a weighted empirical distribution is estimated for each source at each point in time, ambiguity balls centred around these empiricals, and the result intersected. Depending on the choice of the radii, it is clear that this subsumes our weighted approach. Moreover, such a hybrid approach would be able to make use of all of the historical samples, some of which are otherwise weighted to zero and essentially ignored by our weighted approach. Although the intersection (and thus this hybrid) approach is computationally expensive for a large number of sources, it might be possible to keep it manageable by using, say, only two or three constituent-source ambiguity sets.

Chapter 6

Nonstationary Distribution Estimation

In the previous chapter we considered how to estimate ambiguity sets from nonstationary data using Wasserstein balls centred at weighted empirical distributions. Our analysis was guided by concentration bounds and carried out on average, in the sense that we identified *fixed* weights which yielded probabilistic out-of-sample guarantees under the data-sampling distributions. The resulting weights were monotonic in the age of the historical samples and had a simple truncated-polynomial form. However, this neglects the particular values of the historical samples, which may reveal additional information that could be used to adjust the weights for each sample *per* data outcome and hence improve performance.

In this chapter we again study the estimation of probability distributions from nonstationary data, but with the weights informed by the particular data outcome. We suppose that historical observations are generated from a time-evolving distribution with the likelihood of a shift from one distribution to another decreasing with the Wasserstein distance between them. This induces a likelihood depending on both the age of the historical samples *and* their values. We hence take a maximum-likelihood approach to the estimation of nonstationary distributions, and aim to apply this to forecasting the dairy prices from Chapter 4.

This chapter is based on the joint work:

> **Nonstationary Distribution Estimation via Wasserstein Probability Flows**
> *Edward J. Anderson & Dominic S. T. Keehan* [5].

Before detailing our approach, we provide a broad overview of the literature on estimation and optimization with nonstationary data. Nonstationarity may arise in different ways: one approach models the underlying distribution as changing a limited number of times. A typical framework allows at most $O(\log(T))$ changes among T historical observations, so that as the history grows, there are more opportunities to learn the current distribution before another change occurs [27].

Nonstationary versions of the classical newsvendor problem have been studied extensively. One class of models permits changes to the distribution's mean, constrained by a fixed budget of variation. Variation is measured by taking the maximum over any time-epoch subsequence of the sum of squared distances between consecutive means (akin to a sum of squared peak-to-trough distances) [72, 3]. Similar formulations are also used for pricing problems where sellers face nonstationary demand [73].

An alternative approach to modelling nonstationarity assumes that the distribution is governed by an underlying state which evolves according to a Markov chain. This gives rise to partially observable Markov decision processes, as studied in the inventory setting of [117]. Nonstationarity in the transition dynamics themselves can also be captured by imposing variation bounds on the transition probabilities, as in [28].

The fundamental problem of estimating a nonstationary distribution is often approached parametrically. For example, when the underlying distribution is known to be normal and its mean evolves over time according to a Gaussian process, the Kalman filter provides a natural estimation framework. Indeed, the Kalman filter can also be applied without strict normality assumptions [65], but this estimates a latent state with noisy dynamics and measurements, whereas our aim is to estimate the distribution itself directly from the observed data.

This problem can also be viewed from the perspective of time-series forecasting, where the distribution of the next period's observation is effectively estimated from the forecast mean and variance (a point forecast and an associated confidence interval). Nonstationarity is reflected in the fact that both the mean and variance may change over time. Many forecasting models incorporate error terms that exhibit heteroscedasticity — this is the basis of the celebrated ARCH and GARCH models, which have proven highly effective in modelling financial time series. But these models differ from ours in that the error term with changing variance relates to the innovation. There are many versions of these models [23]; typically the time-series value is given by an autoregressive term $\boldsymbol{y}_t = a_0 + \sum_{s=1}^{t-1} a_s y_s + \boldsymbol{\varepsilon}_t$ in which $\boldsymbol{\varepsilon}_t$ is drawn from a distribution with mean 0 and variance σ_t^2, and the variance depends on the previous values

of both ε_t and σ_t^2. In an ARCH model, $\boldsymbol{y}_t$ represents the actual value at time t, whereas our model combines a random innovation term in the underlying mean value $\boldsymbol{y}_t$ with a measurement error, giving an observation $\boldsymbol{\xi}_t = \boldsymbol{y}_t + \boldsymbol{\eta}_t$, where the distribution of the measurement error $\boldsymbol{\eta}_t$ can vary over time as well.

Two simple and widely used approaches for dealing with evolving data are the *rolling window*, where estimation is based solely on the most recent observations, and its generalisation, *weighted estimation*, where more recent observations are given higher weight [72]. These approaches can also be adjusted dynamically to reflect the stability of the observations [64].

When varying weights are applied to observations, an appealing choice is exponential decay, as used in simple exponential smoothing for time-series forecasting. Exponential smoothing is optimal in a variety of settings; see [26] for a comprehensive review. For example, simple exponential smoothing provides the optimal forecast when the observation $\boldsymbol{\xi}_t$ is given by an underlying level $\boldsymbol{y}_t$ plus a noise $\boldsymbol{\varepsilon}_t$, with the $\boldsymbol{y}_t$ subject to random innovations $\boldsymbol{\delta}_t$ from one period to the next, a correlation between $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$, and successive values of both the noise and innovation each independent and identically distributed.

In standard forecasting applications, exponential smoothing generates a single-point forecast. In contrast, our approach applies exponentially decaying weights to the empirical distribution of observations, producing a new (still discrete) weighted empirical distribution, which serves as an estimate of the evolving underlying distribution. For example, in a nonstationary newsvendor setting we would choose an order quantity based on the appropriate quantile of this unequally weighted empirical distribution. Similar ideas are explored in [25, 2].

Our approach to nonstationary distribution estimation is fully nonparametric. To model changes in the distribution over time, we introduce a measure of temporal variation using the Wasserstein distance. Specifically, starting from a given distribution, we assume that the likelihood of transitioning to a new distribution is determined by the Wasserstein distance to that distribution, with the probability decreasing as the distance increases.

In the case of independent samples drawn from a *fixed* distribution, the maximum-likelihood estimator is the empirical distribution that assigns equal weight to each sample. Recall that this gives the widely used sample average approximation (or SAA) approach to data-driven stochastic optimization, where each historical observation is treated equally. Thus, SAA can be interpreted as a maximum-likelihood procedure under the assumption of distributional stationarity.

The SAA approach is nonparametric and is effective even in circumstances when the true underlying distribution is known to be continuous rather than discrete. One might imagine that additional knowledge about the true distribution (for example Lipschitzian properties of the density) could be used to improve the stochastic optimization procedure, but SAA performs very well even in comparison to much more complex approaches [4].

When the underlying distribution evolves over time, we propose an extension of the maximum-likelihood framework that penalises large changes in the estimated distributions between consecutive time periods. As this shift penalty approaches infinity, the method reduces to standard SAA, lending support to its effectiveness even in more general settings where additional distributional structure may be known.

The shift penalty we propose is proportional to the 1-Wasserstein distance between consecutive distributions. Because of the triangle inequality, this gives rise to a natural interpretation in terms of flow through a network. This insight forms the basis of our proposed method, which we term *Wasserstein Probability Flow* (*or* WPF). Unlike other measures of divergence — such as the Kullback–Leibler divergence, the Wasserstein distance leads to a tractable optimization problem that can be efficiently solved using standard network flow techniques. Moreover, it also incorporates geometric information contained in the distances between each of the historical observations (often neglected by other divergence-based approaches), and gives the user flexibility to vary the metric to suit a particular application.

We make three contributions in this chapter:

(i) **Introduce the WPF method.** We introduce the WPF method for estimating probability distributions from nonstationary time series. To the best of our knowledge, this approach is novel and has not appeared in the existing literature.

(ii) **Establish properties of the WPF solution.** In particular, we show that the resulting distribution is uniquely determined under mild conditions and (effectively) supported only on the observed data points.

(iii) **Empirically evaluate the WPF method.** We demonstrate through numerical testing that the WPF method performs well on both synthetic and real-world data sets. Notably, we apply it to the price-forecasting component of the dairy supply chain problem from Chapter 4 and improve upon SAA and other weight-based approaches.

The remainder of the chapter is organised as follows. In Section 6.1 we present the model underlying the WPF method and show that in each time period the distribution we find at optimality is discrete. Section 6.2 gives the network-flow formulation. Finally in Section 6.3 we report on numerical tests of the performance of the WPF method, and Section 6.4 concludes with a discussion.

6.1 Model

In this section we give details on our model setup and introduce the WPF method. Throughout the chapter we will work only with the *1*-Wasserstein distance, which we refer to as and denote simply by *the* Wasserstein distance and W. Recall that for a closed outcome space $\Xi \subseteq \mathbb{R}^m$ and a Polish metric space (Ξ, d), the *Wasserstein distance* between two distributions $\mathbb{P}, \mathbb{Q} \in \mathfrak{P}(\Xi)$ is

$$W(\mathbb{P}, \mathbb{Q}) \coloneqq \inf_{\gamma \in \Gamma(\mathbb{P}, \mathbb{Q})} \int_{\Xi \times \Xi} d(\xi, \zeta)\, d\gamma(\xi, \zeta),$$

where $\Gamma(\mathbb{P}, \mathbb{Q})$ is the set of all couplings of $\mathbb{P}$ and $\mathbb{Q}$; that is, the set of all probability distributions on $\Xi \times \Xi$ with first marginal $\mathbb{P}$ and second marginal $\mathbb{Q}$. As assumed throughout, the distance d is lower semicontinuous which implies that the infimum is attained [118, Theorem 4.1].

Suppose that there is a sequence of underlying distributions $\{\mathbb{P}_t \in \mathfrak{P}(\Xi)\}_{t \in [T]}$ generated from some unknown stochastic process. For some $\lambda \geq 0$ we set the likelihood of a shift from $\mathbb{P}_t$ to $\mathbb{P}_{t+1}$ proportional to

$$\exp\big(-\lambda W(\mathbb{P}_t, \mathbb{P}_{t+1})\big).$$

Under the assumption that shifts between periods are independent of the history of previous distributions, the likelihood of realising the sequence $\{\mathbb{P}_t\}_{t \in [T]}$ given $\mathbb{P}_1$ is proportional to

$$\prod_{t=1}^{T-1} \exp\big(-\lambda W(\mathbb{P}_t, \mathbb{P}_{t+1})\big).$$

Our discussion here is informal. We deal with likelihoods and so the relative probabilities are the key quantities. But to do this in a formal sense would require the equivalent of a density on the infinite-dimensional space of T-tuples of distributions on Ξ, and this is beyond the scope of this chapter.

Randomly sampling an observation from each of the distributions $\mathbb{P}_1, \ldots, \mathbb{P}_T$, the likelihood of observing the outcomes $\xi_1, \ldots, \xi_T \in \Xi$ is the product $\prod_{t=1}^{T} \mathbb{P}_t(\{\xi_t\})$. Hence, the likelihood of realising the sequence $\{\mathbb{P}_t\}_{t\in[T]}$ given $\{\xi_t\}_{t\in[T]}$ is proportional to

$$\prod_{t=1}^{T} \mathbb{P}_t(\{\xi_t\}) \cdot \prod_{t=1}^{T-1} \exp\bigl(-\lambda W(\mathbb{P}_t, \mathbb{P}_{t+1})\bigr).$$

Thus, for observations $\widehat{\xi}_1, \ldots, \widehat{\xi}_T \in \Xi$, we solve the problem

$$\underset{\mathbb{P}_1,\ldots,\mathbb{P}_T \in \mathfrak{P}(\Xi)}{\text{maximize}} \quad \sum_{t=1}^{T} \log\bigl(\mathbb{P}_t(\{\widehat{\xi}_t\})\bigr) - \lambda \cdot \sum_{t=1}^{T-1} W(\mathbb{P}_t, \mathbb{P}_{t+1}) \tag{WPF}$$

to obtain a maximum-likelihood estimate for the sequence of underlying distributions. This has the natural interpretation of simply maximizing log-likelihood with a penalty on changes in the distribution between time periods. Note that we may have $\mathbb{P}_t(\{\widehat{\xi}_t\}) = 0$, so in order for (WPF) to be well defined, throughout the chapter we set $\log(0) \coloneqq -\infty$ and $\exp(-\infty) \coloneqq 0$.

Remark 6.1. Large values of λ correspond to a large penalty on changes in the distribution between periods. If $\lambda = \infty$, then every distribution in the optimal solution to (WPF) is the empirical distribution on the observations. In contrast, if $\lambda = 0$, then there is no penalty on changes in the distribution between periods and the sequence of point-mass distributions on each observation is optimal.

Through the definition of the Wasserstein distance, (WPF) has minimizations in its objective function. These terms are multiplied by $-\lambda \leq 0$, and thus we include them in the overall maximization. For a feasible solution to (WPF) with distributions $\mathbb{P}_1, \ldots, \mathbb{P}_T$, since the infimum in the definition of the Wasserstein distance is attained (as assumed throughout), there exist consecutive minimal-cost transport plans $\gamma_1, \ldots, \gamma_{T-1} \in \mathfrak{P}(\Xi^2)$ between $\mathbb{P}_1, \ldots, \mathbb{P}_T$. Applying the gluing lemma [118, page 23], it follows that a distribution $\mathbb{P} \in \mathfrak{P}(\Xi^T)$ with associated marginals $\mathbb{P}_1, \ldots, \mathbb{P}_T$ and joint marginals $\gamma_1, \ldots, \gamma_{T-1}$ exists. Expressing $\log\bigl(\mathbb{P}_t(\{\widehat{\xi}_t\})\bigr)$ as $\log\bigl(\int_{\Xi^T} \mathbb{1}\{\xi_t = \widehat{\xi}_t\}\, d\mathbb{P}(\xi_1, \ldots, \xi_T)\bigr)$ and $W(\mathbb{P}_t, \mathbb{P}_{t+1})$ as $\int_{\Xi^T} d(\xi_t, \xi_{t+1})\, d\mathbb{P}(\xi_1, \ldots, \xi_T)$ enables us to reformulate (WPF) as

$$\underset{\mathbb{P}\in\mathfrak{P}(\Xi^T)}{\text{maximize}} \quad \sum_{t=1}^{T} \log\biggl(\int_{\Xi^T} \mathbb{1}\{\xi_t = \widehat{\xi}_t\}\, d\mathbb{P}(\xi_1, \ldots, \xi_T)\biggr) - \lambda \cdot \sum_{t=1}^{T-1} \int_{\Xi^T} d(\xi_t, \xi_{t+1})\, d\mathbb{P}(\xi_1, \ldots, \xi_T). \tag{6.1}$$

Here the distributions $\mathbb{P}_t$ in the original formulation are simply the marginals of $\mathbb{P}$.

6.1.1 Finite-Dimensional Reduction

Optimizing over arbitrary probability distributions in $\mathfrak{P}(\Xi)$ is an infinite-dimensional problem. Our first result relies on the lower semicontinuity of the distance d through the reformulation (6.1), and shows that no suboptimality is imposed by instead optimizing over distributions supported only at a finite number of points. Let $\mathfrak{P}_\nu(\Xi) \subseteq \mathfrak{P}(\Xi)$ denote the set of discrete probability distributions supported on at most ν points within Ξ.

Lemma 6.1. *For any feasible solution to (WPF), there exists another feasible solution consisting of discrete distributions each supported on at most $T+1$ points in Ξ, with the same objective value.*

Proof. Let $(\mathbb{P}_1, \ldots, \mathbb{P}_T)$ be a feasible solution to (WPF) and let $\mathbb{P}$ be a feasible solution to (6.1) with marginals $\mathbb{P}_1, \ldots, \mathbb{P}_T$. For this solution $\mathbb{P}$ the objective function of (6.1) features T log terms and we suppose that these take values $c_1, \ldots, c_T \in \mathbb{R} \cup \{-\infty\}$, so $c_t = \log(\int_{\Xi^T} \mathbb{1}\{\xi_t = \widehat{\xi}_t\}\, d\mathbb{P})$.

Replacing each log term in the objective by an equality constraint on its value and exponentiating, we obtain an equivalent linear program over probability measures $\mathbb{Q} \in \mathfrak{P}(\Xi^T)$:

$$\begin{array}{ll} \underset{\mathbb{Q} \in \mathfrak{P}(\Xi^T)}{\text{maximize}} & -\lambda \displaystyle\sum_{t=1}^{T-1} \int_{\Xi^T} d(\xi_t, \xi_{t+1})\, d\mathbb{Q}(\xi_1, \ldots, \xi_T), \\ \text{subject to} & \displaystyle\int_{\Xi^T} \mathbb{1}\{\xi_t = \widehat{\xi}_t\}\, d\mathbb{Q}(\xi_1, \ldots, \xi_T) = \exp(c_t), \quad t \in [T]. \end{array} \tag{6.2}$$

Problem (6.2) is an infinite-dimensional linear program with T equality constraints in the space of Borel probability measures on Ξ^T. By a standard finite-support result for such linear programs (see, e.g., [92] or its application in [119, Appendix B]), there exists an optimal solution $\mathbb{Q}^\star$ supported on at most $T+1$ points in Ξ^T. The marginals of $\mathbb{Q}^\star$ then define discrete distributions $\mathbb{P}_1^\star, \ldots, \mathbb{P}_T^\star$ that are feasible for (WPF) and attain the same objective value as $(\mathbb{P}_1, \ldots, \mathbb{P}_T)$. □

Our next result makes use of the fact that the distance d satisfies the triangle inequality and shows that no suboptimality is imposed by instead optimizing over distributions supported only on points within the set of observations. The result is established by taking probability mass that is not at an observed point and moving it to an observed point, with the triangle inequality ensuring that the total transport costs do not increase.

Proposition 6.1. *For any feasible solution to (WPF), there exists a solution consisting of discrete distributions supported only on points within the set of observations with either the same or an improved objective value.*

Proof. By Lemma 6.1, we may restrict attention to discretely supported solutions of (6.1). Consider such a solution $\mathbb{P} \in \mathfrak{P}_{T+1}(\Xi^T)$ having $\mathbb{P}(\{(\xi_1, \ldots, \xi_T)\}) > 0$ for some point $(\xi_1, \ldots, \xi_T) \in \Xi^T$ with $\xi_t \notin \{\widehat{\xi}_1, \ldots, \widehat{\xi}_T\}$, where the observations are $\widehat{\xi}_1, \ldots, \widehat{\xi}_T$. This assignment of probability mass does not contribute to the t-th log term in the objective function of (6.1), so changing ξ_t only affects the objective value through the distance terms. We consider two types of changes. First suppose that $\xi_{t+1} \in \{\widehat{\xi}_1, \ldots, \widehat{\xi}_T\}$. Then the sum of the distance terms in the objective function would not be increased by instead assigning the probability $\mathbb{P}(\{(\xi_1, \ldots, \xi_T)\})$ to the alternate point $(\xi_1, \ldots, \xi_{t-1}, \xi_{t+1}, \xi_{t+1}, \ldots, \xi_T) \in \Xi^T$, since by the triangle inequality,

$$d(\xi_{t-1}, \xi_{t+1}) \leq d(\xi_{t-1}, \xi_t) + d(\xi_t, \xi_{t+1}).$$

For the second type of change we suppose that $\xi_{t-1} \in \{\widehat{\xi}_1, \ldots, \widehat{\xi}_T\}$. Then we can assign the probability $\mathbb{P}(\{(\xi_1, \ldots, \xi_T)\})$ to the alternate point $(\xi_1, \ldots, \xi_{t-1}, \xi_{t-1}, \xi_{t+1}, \ldots, \xi_T) \in \Xi^T$, also without increasing the total transport costs.

Provided that there is at least one element ξ_s of $(\xi_1, \ldots, \xi_T)$ with $\xi_s \in \{\widehat{\xi}_1, \ldots, \widehat{\xi}_T\}$, then repeatedly making the first type of change for the largest $t < s$ such that ξ_t is not within the observations, or making the second type of change for the smallest $t > s$ such that ξ_t is not within the observations, will produce a solution supported only on the observations which is at least as good as $(\xi_1, \ldots, \xi_T)$. If on the other hand $\xi_1, \ldots, \xi_T \notin \{\widehat{\xi}_1, \ldots, \widehat{\xi}_T\}$, then this point does not contribute positively to the objective function, and the probability $\mathbb{P}(\{(\xi_1, \ldots, \xi_T)\})$ can be assigned to $(\widehat{\xi}_1, \ldots, \widehat{\xi}_1) \in \Xi^T$, thereby improving the objective value.

Hence, all probability mass can be concentrated on tuples whose components belong to the set of observations without worsening the objective value. □

6.2 Network-Flow Reformulation

Building on the finite-dimensional reduction of Proposition 6.1, in this section we show how (WPF) can be reformulated as a network-flow problem. With Proposition 6.1 in hand, noting that the Wasserstein distance between two discrete distributions can be found through a transport linear program, we write $p_t(i)$ for the amount of probability mass assigned to observation i by $\mathbb{P}_t$; $\gamma_t(i, j)$ for the amount of probability mass transported between observations i and j during the shift from $\mathbb{P}_t$ to $\mathbb{P}_{t+1}$; and $d(i, j)$ for the distance between observations i and j. Thus, to find an

optimal solution to (WPF), we solve the problem:

$$\begin{aligned}
\underset{\substack{p_1,\ldots,p_T \\ \gamma_1,\ldots,\gamma_{T-1}}}{\text{maximize}} \quad & \sum_{t=1}^{T} \log(p_t(t)) - \lambda \sum_{t=1}^{T-1} \sum_{j=1}^{T} \sum_{i=1}^{T} d(i,j)\gamma_t(i,j), \qquad (6.3) \\
\text{subject to} \quad & \sum_{i=1}^{T} p_1(i) = 1, \\
& p_t(i) = \sum_{j=1}^{T} \gamma_t(i,j), \qquad i \in [T],\ t \in [T-1], \\
& \sum_{i=1}^{T} \gamma_t(i,j) = p_{t+1}(j), \qquad j \in [T],\ t \in [T-1], \\
& p_t(i) \in \mathbb{R}_+, \qquad i \in [T],\ t \in [T], \\
& \gamma_t(i,j) \in \mathbb{R}_+, \qquad i,j \in [T],\ t \in [T-1].
\end{aligned}$$

This is a type of convex network-flow problem. The probability mass moving through the network is conserved because of the constraints and so the requirement that the probabilities sum to 1 in the first period is retained for all periods. It can be seen that this problem has an optimal solution due to the extreme-value theorem, and we thus conclude that (WPF) has one as well.

In this network, flow is between an observation i at time t and an observation j at time $t+1$, i.e., on arcs between (t,i) and $(t+1,j)$. There is conservation of flow at all nodes (t,i) except the sources and sinks at times $t=1$ and T. We define a *path* as a sequence of nodes at consecutive time intervals $(t,i_0),(t+1,i_1),\ldots,(t+k,i_k)$. Observe that the flow in the network does not contain any cycles, so from the flow decomposition property [1], it can be decomposed into flows along paths (a *path flow*) from nodes at time $t=1$ to time $t=T$. We use this property to prove the result below which restricts the types of flow considered at optimality. Since we are interested in estimation applications which use the terminal distribution, a key property is that the restrictions we make do not alter the terminal probabilities at optimality.

Proposition 6.2. *For any optimal solution to (6.3), we can construct a new optimal solution in which p_T is unchanged and in which $\gamma_t(i,j) = 0$ unless either $i=j$ or $i=t$.*

Proof. We prove this result by construction, starting from an optimal solution to (6.3) and noting that if any $p_T(i)$ is changed then there will be a strict improvement in the objective value. For an optimal solution $(p^\star,\gamma^\star)$, let $\mathcal{F}$ be a path-flow decomposition. Consider $f \in \mathcal{F}$. This involves a flow δ along a path $(1,i_1),\ldots,(T,i_T)$ and we write $\mathcal{T} \coloneqq \{t : i_t = t\}$ for the set of times when this flow contributes to the log terms. (We order the elements $\{t_1,\ldots,t_S\}$ with $t_1 < \cdots < t_S$.)

We now replace the flow f with a new flow f' satisfying the statement of the proposition. For all $t_i < t < t_{i+1}$, we set $f'\big((t, t_{i+1}), (t+1, t_{i+1})\big) = \delta$ and the other f' values to 0 for arcs from t to $t+1$. For all $t = t_i$, we set $f'\big((t, t), (t+1, t_{i+1})\big) = \delta$ and the other f' values to 0 for arcs from t to $t+1$. If $t_1 > 1$, then for all $t < t_1$, we set $f'\big((t, t_1), (t+1, t_1)\big) = \delta$ and the other f' values to 0 for arcs from t to $t+1$. If $t_S < T$, then for all $t_S \leq t < T$ we set $f'\big((t, t_S), (t+1, t_S)\big) = \delta$ and the other f' values to 0 for arcs from t to $t+1$. This change is illustrated in Figure 6.1.

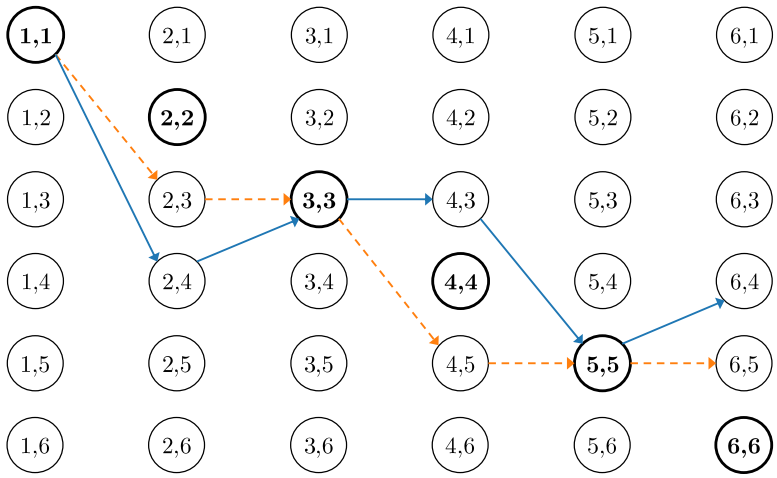

Figure 6.1: **Construction Used in the Proof of Proposition 6.2.** Solid arrows show the initial flow f and dashed arrows show the newly constructed flow f'. Bold nodes on the diagonal contribute to the log terms in the objective, so $\mathcal{T} = \{1, 3, 5\}$ here.

Note that f' is a path flow that matches the flow into the nodes where f contributes to the log terms in the objective. Using the triangle inequality, we can also see that the total distance along f' is no greater than the distance along f, since wherever possible f' moves directly to the next node, rather than possibly moving via other nodes. Moreover, there is zero distance along the sections of the path before t_1 or after t_S.

Repeating this construction for each $f \in \mathcal{F}$ and combining the resulting path flows yields a feasible solution with a no-worse objective value, and in which the only non-negative flows $\gamma_t(i, j) > 0$ occur when $i = j$ or $i = t$. If $p_T^\star$ were altered through any of the changes from f to f', then we must have some $t_S < T$ and $i_T \neq t_S$, in which case there will be nonzero distances in the last section of the path replaced with zero distances. This gives an improvement from the change and therefore contradicts the optimality of $(p^\star, \gamma^\star)$. We conclude that if the original flow is optimal, then so is the new flow. □

This result enables us to rewrite problem (6.3) with a reduced set of flow variables from one observation to a later observation. We consider source and sink nodes 0 and $T+1$, and define a

variable $x(i,j)$ for the probability flow between different nodes i and j. We set

$$\begin{aligned}
&x(0,j) = p_1(j), \qquad j \in [T],\\
&x(i,j) = \gamma_i(i,j), \qquad i \in [j-1],\ j \in [T],\\
&x(i,T+1) = p_T(i), \qquad i \in [T],
\end{aligned}$$

and all other $x(i,j)$ variables to 0, applying Proposition 6.2. As a consequence,

$$\sum_{i=0}^{T} x(i,j) = p_j(j), \quad j \in [T].$$

Thus, to find the terminal distributions in optimal solutions to (WPF), we can solve the problem:

$$\begin{aligned}
\underset{x}{\text{maximize}} \quad & \sum_{j=1}^{T} \log\Big(\sum_{i=0}^{T} x(i,j)\Big) - \lambda \cdot \sum_{j=1}^{T}\sum_{i=1}^{T} d(i,j)x(i,j), && (6.4)\\
\text{subject to} \quad & \sum_{j=1}^{T} x(0,j) = 1,\\
& \sum_{i=0}^{T} x(i,j) = \sum_{k=1}^{T+1} x(j,k), \qquad j \in [T],\\
& x(i,j) = 0, \qquad j \in [i],\ i \in [T],\\
& x(0,T+1) = 0,\\
& x(i,j) \in \mathbb{R}_+, \qquad i \in \{0\} \cup [T],\ j \in [T+1].
\end{aligned}$$

Ignoring variables which are fixed to 0, the reduced problem (6.4) has $O(T^2)$ variables and $O(T)$ constraints, whereas the original problem (6.3) has $O(T^3)$ variables and $O(T^2)$ constraints.

We may reformulate (6.4) by replacing each $\log(\sum_{i=0}^{T} x(i,j))$ term in the objective with a variable $z_j \in \mathbb{R}$ and imposing the constraint $z_j \leq \log(\sum_{i=0}^{T} x(i,j))$. These amount to exponential-cone constraints, and the result is an exponential-cone program that can be readily addressed with off-the-shelf solvers.

The following result ensures that (6.4) has unique $x(i,T+1)$, $i \in [T]$, values at optimality, thus implying that the optimal terminal distribution for (6.3) is unique, with the same implication for (WPF) when restricting to discrete distributions supported only on the set of observations. We say that a set of scalars $\{c_1,\ldots,c_N\}$ has *unique subset sums* if there are no two subsets $\mathcal{I},\mathcal{I}' \subseteq [N]$ with $\mathcal{I} \neq \mathcal{I}'$ and $\sum_{i\in\mathcal{I}} c_i = \sum_{i\in\mathcal{I}'} c_i$.

Theorem 6.1. *If each set of the nonzero distances along any acyclic path between two different observations (i.e., a set $\{d(i_1, i_2), d(i_2, i_3), \ldots, d(i_{k-1}, i_k)\}$ with $i_1, i_2, \ldots, i_k$ all differing) has unique subset sums, then every optimal solution to problem (6.3) has the same terminal probability distribution p_T; that is, the optimal terminal distribution is unique.*

Proof. We prove the result for problem (6.4) which establishes the result for problem (6.3) through the equivalence provided by Proposition 6.2. Problem (6.4) is a convex optimization problem, so the optimal solutions form a convex set. Consider two optimal solutions x and x' that differ on the flows to $T+1$; since $x(i, T+1) = p_T(i)$, obtaining a contradiction will be enough to establish the result.

Moving along the line of convex combinations of x and x', strict concavity of log implies that $\sum_{i=0}^{T} x(i,j) = \sum_{i=0}^{T} x'(i,j)$ for each $j \in [T]$. Thus the objective values for x and x' must match for the log terms in the objective. Since these are both optimal solutions and attain the same objective value, we deduce for the distance terms in the objective that

$$\sum_{j=1}^{T}\sum_{i=1}^{T} d(i,j)x(i,j) = \sum_{j=1}^{T}\sum_{i=1}^{T} d(i,j)x'(i,j).$$

Now consider $x - x'$: this is a nonzero flow through the network with positive and negative components and zero net flow in and out of all nodes (including 0 and $T+1$). It can hence be decomposed into cycles. By assumption, $x - x'$ includes a nonzero flow to $T+1$. Among these cycles including $T+1$, we choose x'' and write i_1 and i_2 for the two nodes adjacent to $T+1$.

Note that for any pair (i,j), we have that $x''(i,j)$ is nonzero only if $x(i,j) - x'(i,j)$ is nonzero, and hence at least one of $x(i,j)$ or $x'(i,j)$ is nonzero. Now consider solutions $(x+x')/2 \pm \delta x''$. Because x'' is a cycle these two solutions satisfy conservation of flow at every node other than 0 and $T+1$, and their components are all non-negative for small enough δ. Thus they are both feasible solutions to problem (6.4).

Since x and x' are both optimal, we arrive at a contradiction unless

$$\sum_{j=1}^{T}\sum_{i=1}^{T} d(i,j)x''(i,j) = 0. \tag{6.5}$$

(We exclude $x''(i_1, T+1)$ and $x''(T+1, i_2)$ from this sum as the multiplying distance is 0.) With x'' being a cycle, each nonzero component has the same magnitude. In view of (6.5), splitting x'' into positive and negative elements then contradicts the unique subset sum property

for the path from i_1 to i_2 that does not visit $T+1$. Hence, our assumption that x and x' differ on the flows to $T+1$ was wrong and the result is established. □

The unique subset sum condition of Theorem 6.1 is unrestrictive when Ξ is not discrete. In this case the condition will always hold for a small perturbation of the observations.

In the case where the set of all nonzero distances has the unique subset sum property (which is a stronger condition than that of Theorem 6.1) we have the stronger result that there is a unique optimal solution to (6.4). However, this condition tends to fail when Ξ is one dimensional, due to the fact that transporting mass from observation $\widehat{\xi}_1$ to observation $\widehat{\xi}_2$ has the same cost as going via $\widehat{\xi}_3$ if $\widehat{\xi}_3$ lies between $\widehat{\xi}_1$ and $\widehat{\xi}_2$.

Example 6.1. It will be helpful to consider a small one-dimensional example. Suppose that we have the 6 observations 6.13, 7.85, 6.47, 4.91, 5.54, and 7.13 (in sequence), and we use the standard Euclidean norm to compute distances. Solving (6.4) with $\lambda = 4$ yields the optimal objective value -8.7052 and a set of flows shown in Figure 6.2.

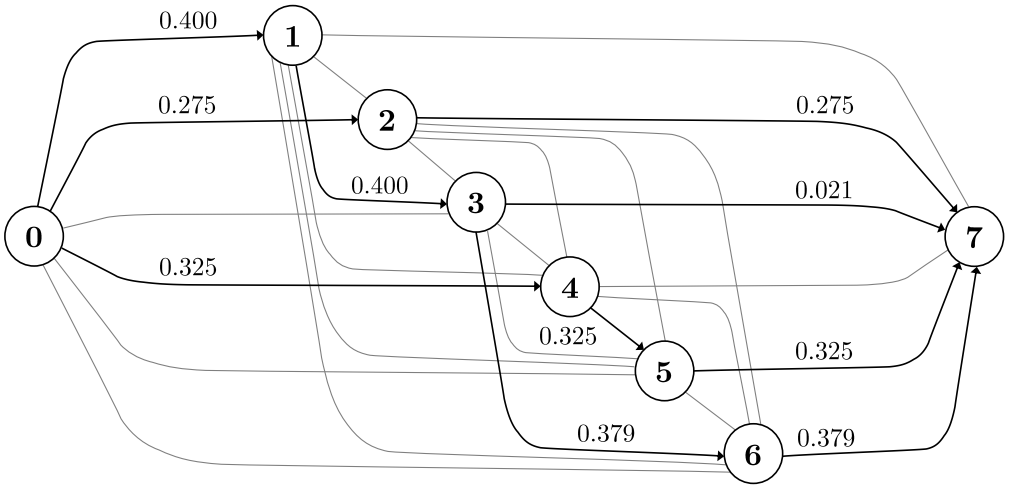

Figure 6.2: **Optimal Solution for Example 6.1.** Arcs that have no flow are shown in grey.

The terminal distribution assigns probability 0.275 to observation 2 (7.85), probability 0.021 to observation 3 (6.47), probability 0.325 to observation 5 (5.54), and probability 0.379 to observation 6 (7.13). Note that reversing time yields the same optimal flow pattern with the directions reversed. Thus, we can also see from the figure the optimal solution to (6.4) if the observations were instead collected in reverse order. The terminal distribution then assigns probabilities 0.400, 0.275 and 0.325 to observations 6.13, 7.85 and 4.91 respectively. This demonstrates that the order that the observations are collected in influences the optimal solution for the WPF method. □

6.3 Numerical Results

In this section we evaluate the performance of the WPF method. We begin by demonstrating that the WPF method has value over other weighting approaches using a synthetic newsvendor problem. This setup allows us to estimate expected performances with high statistical precision. Our findings show that the dimensionality and complexity of the underlying distributions significantly influence how beneficial (or detrimental) WPF can be. We then test performance on the price-forecasting component of the dairy supply chain problem from Chapter 4 and a portfolio-optimization application. Here observations are taken from multidimensional real-world data sets collected over long timescales, and we use a parameter-tuning scheme which could be feasibly employed in practice.

As benchmarks, we compare to the performance of SAA, windowing, and simple exponential smoothing. Like WPF, these all define weighted empirical distributions on the observations. SAA sets

$$p_T(1) = 1/T,\ \ldots,\ p_T(T) = 1/T,$$

windowing for window size $s \in [T]$ sets

$$p_T(T) = 1/s,\ \ldots,\ p_T(T-s+1) = 1/s, \quad p_T(T-s) = 0,\ \ldots,\ p_T(1) = 0,$$

and (simple exponential) smoothing for decay rate $\alpha \in [0, 1]$ sets

$$p_T(T) = \bar{\alpha},\ p_T(T-1) = \bar{\alpha}(1-\alpha),\ p_T(T-2) = \bar{\alpha}(1-\alpha)^2,\ \ldots,\ p_T(1) = \bar{\alpha}(1-\alpha)^{T-1}.$$

The value $\bar{\alpha} = \alpha/(1-(1-\alpha)^T)$ is chosen so that the probabilities sum to 1.[1]

Our evaluations of the WPF method use the exponential-cone reformulation of problem (6.4), implemented in `Julia` [19] with `JuMP.jl` [83] and `MathOptInterface.jl` [80], and solved with `COPT` [53]. Distances are measured using the metrics induced by the 1-, 2-, and ∞-norms, which we refer to by L_1, L_2, and L_∞, respectively.

6.3.1 Multi-Modal and Multi-Dimensional Newsvendor

Consider a newsvendor dealing in m goods, having underage and overage cost vectors $c_\mathrm{u}, c_\mathrm{o} \in \mathbb{R}^m_+$, supplying a vector of random demands $\boldsymbol{\xi} \in \mathbb{R}^m \sim \mathbb{P}$. The vector of optimal order quantities for

[1] For $\alpha = 0$, we take the limiting form as $\alpha \to 0$, corresponding to $p_T(T) = \cdots = p_T(1) = 1/T$.

each good solves

$$\underset{x\in\Xi}{\text{minimize}} \quad \mathbb{E}_{\mathbb{P}}\Big[c_{\text{u}}^{\intercal}(\boldsymbol{\xi}-x)_{+}+c_{\text{o}}^{\intercal}(x-\boldsymbol{\xi})_{+}\Big].$$

For the nonstationary demand process, we use a mixture of n multivariate normal distributions with modes that evolve independently over time. At each time $t \in [T]$ we suppose that

$$\boldsymbol{\xi}_t \in \mathbb{R}^m \sim \frac{1}{n}\sum_{i=1}^{n} \text{Normal}\Big(\mu_t(i), \text{Diag}(\sigma^2,\ldots,\sigma^2)\Big),$$

where $\mu_t(i) \in \mathbb{R}^m$ is the mean vector of mode i at time t, and $\text{Diag}(\sigma^2,\ldots,\sigma^2)$ is an m-by-m diagonal matrix with entries σ^2. For each mode i we further suppose that

$$\boldsymbol{\mu}_{t+1}(i) = \mu_t(i) + \boldsymbol{\varepsilon}_t(i) \quad \text{where} \quad \boldsymbol{\varepsilon}_t(i) \in \mathbb{R}^m \sim \text{Normal}\Big((0,\ldots,0), \text{Diag}(\rho^2,\ldots,\rho^2)\Big).$$

In this way the scalar ρ parametrises the extent of nonstationarity. For simplicity we use isotropic normal distributions and do not vary the covariance matrices over time.

In the case that there is a single mode then we have a problem for which the Kalman filter is appropriate; there is a random walk carried out by a state (the mean $\boldsymbol{\mu}_t$), with observations occurring with a normal measurement error. The usual Kalman filter recursions then give an optimal estimate for the current mean. This can be shown, after some initial transient behaviour, to be given by an exponentially smoothed weighting of the observations. Thus, it is not surprising that smoothing performs very well when there is a single mode, as is shown in Table 6.3 below. There is, however, a significant effect that comes into play when the distribution, rather than the mean, is needed for the stochastic optimization problem of interest. When taking a weighted empirical distribution the mean will be essentially the optimal estimator for the mean of the underlying distribution, but the variance of the weighted empirical distribution will consistently be larger than the variance of the true underlying distribution.

To illustrate the behaviour of the WPF method, Figure 6.3 shows an instance of what happens when there are two modes and we work in a single dimension. This is for one set of realisations of the process and a particular shift penalty.

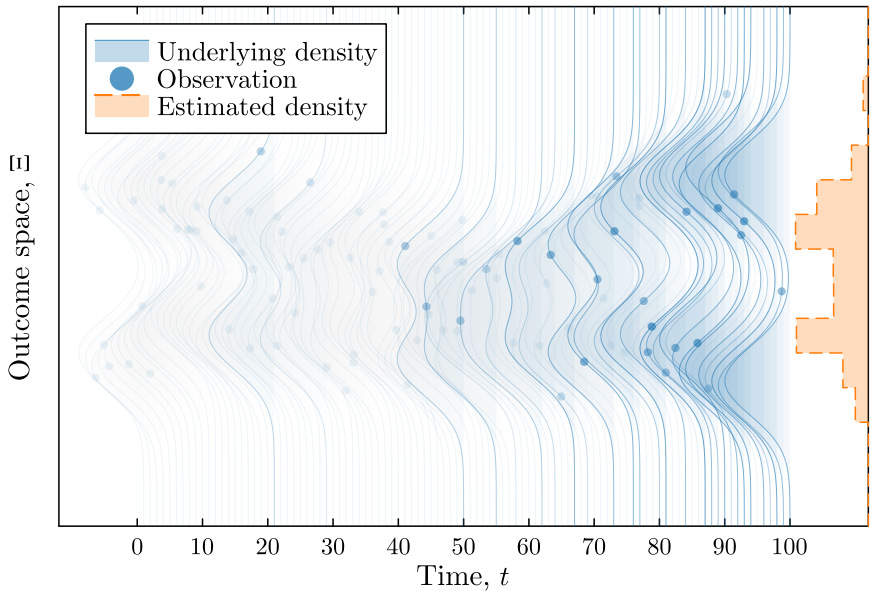

Figure 6.3: **Underlying Distributions, Observations, and Terminal WPF Estimate.** Opacities reflect probabilities in the terminal estimate. Here $\sigma = 20$, $\rho = 4$, and $\lambda = 10$.

Next we consider an example where we use different estimation methods to determine the optimal newsvendor order quantities given observations from the nonstationary process $\boldsymbol{\xi}_t$. In all our examples we have an initial set of modes given by $\mu_1(i) = i(100, \ldots, 100)$, $i \in [n]$, distribution parameters $\sigma = 20$ and $\rho = 15$, and cost vectors $c_{\mathrm{u}} = (4, \ldots, 4)$ and $c_{\mathrm{o}} = (1, \ldots, 1)$. We also use a history length of $T = 100$.

To compare different methods we take the average performance over 1000 realisations and tune the values of the relevant parameters across the set of possible values shown in Table 6.1. Here for scalars $a \leq b$ and a positive integer r, we use $\texttt{LinRange}(a, b; r)$ for the set of r equally spaced values between a and b, inclusive. Similarly for scalars $a \leq b \in (0, \infty)$ and a positive integer r, we use $\texttt{LogRange}(a, b; r)$ for the set of r values forming a geometric sequence between a and b, inclusive. This logarithmic spacing is the natural choice for parametrising windowing and smoothing.

Table 6.1: **Newsvendor-Ordering Parameter Ranges**

Window size	$s \in \lceil \texttt{LogRange}(1, 100; 30) \rceil$
Decay rate	$\alpha \in \{0\} \cup \texttt{LogRange}(10^{-4}, 10^{0}; 30)$
Shift penalty	$\lambda \in \{0\} \bigcup_{i=-3}^{-1} \texttt{LinRange}(10^{i}, 10^{i+1}; 10) \cup \{\infty\}$

First we consider an example with two goods and a demand distribution with three modes. The results are shown in Table 6.2. We use three different distance norms from which to derive

the Wasserstein distances. In this case all do well, with a distance based on the L_2 norm doing best. In the other numerical examples we discuss below, it turns out that switching from e.g. L_1 to L_2 can make a big difference, but here it does not matter so much. We find that both windowing and smoothing give a significant improvement over the basic SAA approach that ignores the nonstationarity. But all versions of the WPF approach are better than these other methods, with cost differences of more than $\approx 3\,\%$ between WPF and Smoothing.

Table 6.2: **Newsvendor-Ordering Performance.** Standard errors ($\pm$ terms) reported in "Difference from SAA" row computed using common random numbers.

	SAA	Windowing	Smoothing	WPF with metric		
				L_1	L_2	L_∞
Ex-post optimal expected cost	427.2	385.6	377.6	368.8	368.0	368.1
Difference from SAA (%)		-9.7 ± 0.5	-11.6 ± 0.5	-13.7 ± 0.5	-13.9 ± 0.5	-13.8 ± 0.5

Next we consider a range of different dimensions m and numbers of modes n for the problem. The results are shown in Table 6.3. We can see that good performance of the WPF method is achieved when there are a greater number of modes and the dimension is greater than one. Roughly speaking, the more complicated the nonstationary behaviour, the better WPF will perform in comparison to other methods.

Table 6.3: **Percentage Cost Difference from Smoothing for WPF with L_1 Metric**

Dimensions	Underlying modes			
	1	2	3	4
1	12.8 ± 1.3	0.2 ± 0.7	-0.9 ± 0.5	0.5 ± 0.3
2	11.7 ± 0.9	-1.0 ± 0.6	-2.3 ± 0.3	-1.0 ± 0.2
3	11.4 ± 0.8	-1.4 ± 0.6	-3.8 ± 0.3	-1.3 ± 0.2
4	11.3 ± 0.7	-2.4 ± 0.6	-3.7 ± 0.3	-1.7 ± 0.2

6.3.2 Price Forecasting for the New Zealand Dairy Industry

In the remainder of this section we adopt a training-and-testing approach to evaluate performance, splitting time-series data into a training set (the first 70 % of the data) and a testing set (the remaining 30 % of the data). In the training phase decisions are not made during a warm-up period while initial observations are collected. After this, decisions are made sequentially, incurring costs based on the subsequent observation from within the training set. These costs inform parameter tuning in the testing phase, where at each time step the method selects the parameter that minimizes the total cost over a fixed parameter tuning window.

In the testing phase (which uses the testing set), decisions continue to be made using all prior observations, with parameters updated dynamically based on performance in the parameter tuning window. This window introduces a trade-off: a longer window yields more stable parameter estimates, while a shorter window responds more quickly to changes in time-series dynamics. In our evaluations we found that a two-year tuning window worked well for WPF, and that varying this window had little effect on the benchmarks. Note that in this training-and-testing setup, decision making and parameter tuning rely solely on data collected in the past, i.e., the tests are consistent with how this method could be employed in practice [115].

Recall that the international market prices of dairy commodities are influenced by factors such as governmental regulations and global demand, both of which evolve over time. Such prices inform the operational decision making of dairy companies around the world, and in Chapter 4 we showed that a model predictive control (MPC) scheme — at the heart of which is a price-forecasting problem — works well for the New Zealand dairy industry. Thus, towards improving the data-driven decision making of the New Zealand dairy industry, we now seek to develop effective price forecasts using the WPF method.

Figure 6.4 graphs fourteen years of monthly Global Dairy Trade (GDT) prices for Anhydrous Milk Fat (AMF), Butter (BUT), Butter Milk Powder (BMP), Skim Milk Powder (SMP), and Whole Milk Powder (WMP). The GDT price is the average per-unit price of each product sold at international commodity auctions in a given month.

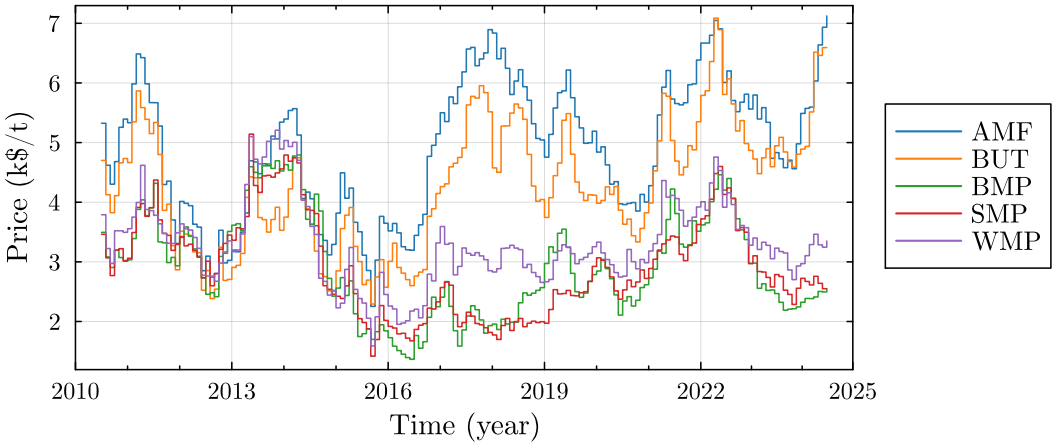

Figure 6.4: **Historical GDT Prices.** Monthly observations; June 2010–May 2024.[2]

Figure 6.4 shows that the GDT prices of each product are correlated. In particular, the prices of the fat products (AMF and BUT) and the milk powders (BMP, SMP, and WMP) are coupled throughout the fourteen-year period, with a change in behaviour around 2016.

To forecast future GDT prices, we use a weighted linear regression on the previous log prices. For historical log prices $\ell_1, \ldots, \ell_T \in \mathbb{R}^m$ observed at times $t = 1, \ldots, T$, assigning the pairs $(\ell_1, \ell_2), \ldots, (\ell_{T-1}, \ell_T)$ probabilities $p_1, \ldots, p_{T-1}$, we forecast the log price at time $T+1$ as $\mu^\star + A^\star \ell_T$, where $\mu^\star$ and $A^\star$ solve

$$\underset{\mu \in \mathbb{R}^m, A \in \mathbb{R}^{m \times m}}{\text{minimize}} \quad \sum_{t=1}^{T-1} p_t \left\| \ell_{t+1} - (\mu + A\ell_t) \right\|_2^2.$$

(Note that this has an analytic solution; see, e.g., [101, Section 11.5].) When the next log price ℓ_{T+1} is realised, we incur the cost $\left\| \ell_{T+1} - (\mu^\star + A^\star \ell_T) \right\|_2^2$. The weighted regression approach here allows adaptation to shifting price dynamics.

Table 6.4 presents the parameter ranges used when tuning, and Table 6.5 presents the price-forecasting performance of the WPF method and the benchmarks. Table 6.5 shows that the WPF method with the L_1 metric provides $\approx 8\,\%$ better forecasts than SAA, while windowing and smoothing perform similarly to SAA.

Table 6.4: **Price-Forecasting Parameter Ranges**

Window size	$s \in \lceil \texttt{LogRange}(10, 14 \cdot 12; 30) \rceil$
Decay rate	$\alpha \in \{0\} \cup \texttt{LogRange}(10^{-4}, 0.9 \cdot 10^{0}; 30)$
Shift penalty	$\lambda \in \bigcup_{i=1}^{3} \texttt{LinRange}(10^i, 10^{i+1}; 10) \cup \{\infty\}$

[2]Data retrieved from [39].

Table 6.5: **Price-Forecasting Performance.** Standard errors (± terms) reported in "Difference from SAA" row computed using common random numbers.

	SAA	Windowing	Smoothing	WPF with metric		
				L_1	L_2	L_∞
Average testing cost	0.0261	0.0277	0.0267	0.0242	0.0280	0.0251
Difference from SAA (%)		5.8 ± 5.7	2.1 ± 10.6	-7.5 ± 5.7	7.0 ± 4.5	-3.9 ± 8.3

Figure 6.5 graphs average costs of the WPF method for different shift penalties, and Figure 6.6 graphs the probabilities assigned to the sequence of historical observations from Figure 6.4 by WPF when price forecasting.

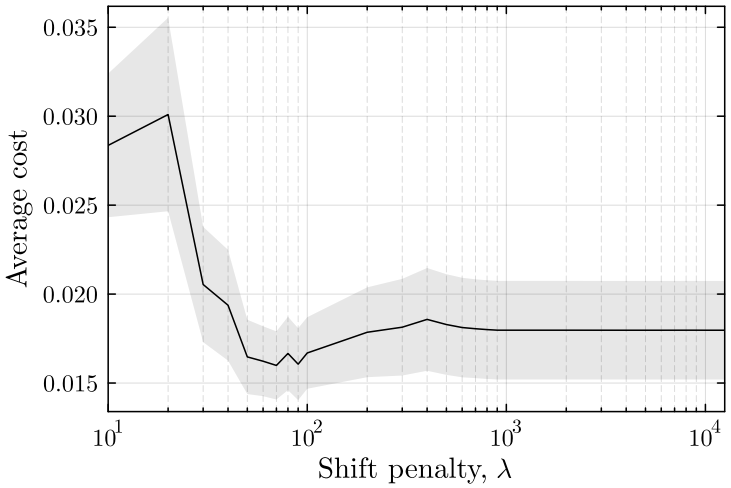

Figure 6.5: **Price-Forecasting Parameter Tuning for WPF with L_1 Metric.** Average cost within the two-year parameter-tuning window. Band presents standard-error range.

Figure 6.5 shows a large decrease in costs as the penalty approaches an intermediate value and the solution begins to utilise the information contained in historical observations, and then a gradual rise and stabilisation in costs as the solution approaches SAA.

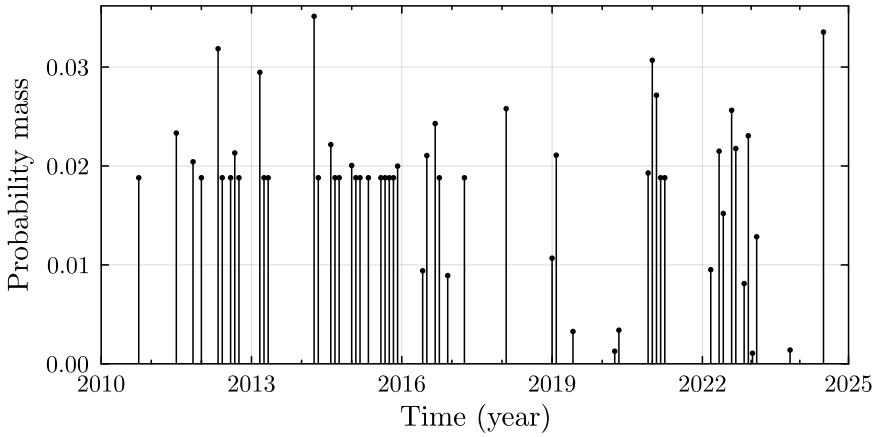

Figure 6.6: **Price-Forecasting Probabilities for WPF with L_1 Metric.** Shift penalty set at the minimum from Figure 6.5.

A careful comparison of Figure 6.6 to Figure 6.4 shows that the WPF method generally only assigns probabilities to observations in which the prices of the two fat products are similar and the prices of each milk-powder product are similar, as is the case for the most recent price dynamics.

In Table 6.5 the SAA approach that we took to forecasting in Chapter 4 is outperformed by WPF (at least with the L_1 metric), and thus we would expect a WPF-augmented MPC scheme to further outperform the proprietary policy if applied in practice. Moreover, this improvement is only possible due to the power of the WPF method, as a windowing- or smoothing-augmented scheme would not be able to selectively utilise information contained in the historical data outcomes, as WPF is able to do in Figure 6.6.

6.3.3 Portfolio Optimization

Figure 6.7 graphs ten years of stock values for Boeing Co. (BA), Berkshire Hathaway Class B (BRK.B), Goldman Sachs Group Inc. (GS), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), Coca-Cola Co. (KO), McDonald's Corp. (MCD), Pfizer Inc. (PFE), Walmart Inc. (WMT), and Exxon Mobil Corp. (XOM).

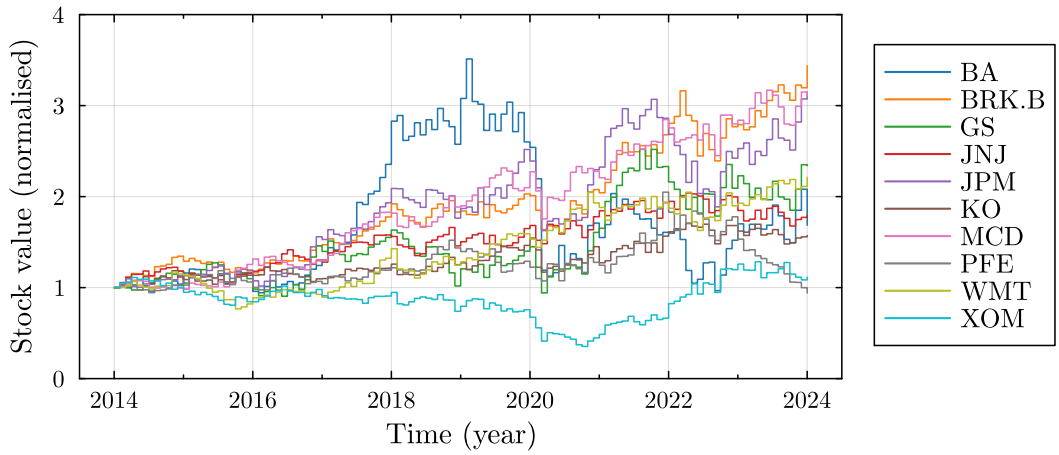

Figure 6.7: **Historical Stock Values.** Monthly observations; Jan. 2014–Dec. 2023.[3]

Figure 6.7 shows that the returns of each stock are correlated, with market shocks affecting most stocks simultaneously. In particular, there are significant shocks in early 2020 and in mid-2022.

We consider a risk-averse portfolio-optimization problem. For m stocks with portfolio weights $x \in \mathbb{R}^m_+$ and random returns $\boldsymbol{\xi} \in \mathbb{R}^m$, the risk-neutral portfolio return is the random variable $x^\intercal \boldsymbol{\xi}$. Adjusting for the conditional value at risk of the portfolio, the problem is

$$\underset{x \in \mathbb{R}^m_+}{\text{minimize}} \;\; (1-\rho) \cdot \mathbb{E}\big[-x^\intercal \boldsymbol{\xi}\big] + \rho \cdot \mathrm{CVaR}_\beta\big[-x^\intercal \boldsymbol{\xi}\big] \;\; \text{subject to} \;\; \sum_{i=1}^m x_i = 1.$$

Here CVaR_β is the conditional-value-at-risk operator at percentile $\beta \in [0,1)$, and $\rho \in [0,1]$ parametrises the level of risk aversion. (Note that β is associated with the worst $100(1-\beta)\,\%$ of returns, so $\beta = 0$ recovers the risk-neutral case.) For T historical returns $\xi_1, \ldots, \xi_T \in \mathbb{R}^m$ assigned probabilities $p_1, \ldots, p_T$, using the Rockafellar–Uryasev formula gives

$$\begin{aligned} \underset{x \in \mathbb{R}^m_+, \tau \in \mathbb{R}}{\text{minimize}} \quad & (1-\rho) \cdot \sum_{t=1}^T p_t\Big(-x^\intercal \xi_t\Big) + \rho \cdot \sum_{t=1}^T p_t\Big(\tau + \tfrac{1}{1-\beta} \cdot \max\{-x^\intercal \xi_t - \tau, 0\}\Big) \\ \text{subject to} \quad & \sum_{i=1}^m x_i = 1. \end{aligned}$$

This can further be expressed as a linear program via an epigraphical reformulation.

In our evaluations we set $\rho = 0.9$ and $\beta = 0.95$. Table 6.6 presents the parameter ranges used when tuning, and Table 6.7 presents the performance of different estimation methods. Table 6.7 shows that the WPF method with the L_1 metric provides $\approx 11\,\%$ better returns than SAA, while windowing and smoothing both perform worse than SAA.

[3] Data retrieved using the `STOCKHISTORY` function in `Microsoft Excel`.

Table 6.6: **Portfolio-Optimization Parameter Ranges**

Window size	$s \in \lceil \texttt{LogRange}(1, 10 \cdot 12; 30) \rceil$
Decay rate	$\alpha \in \{0\} \cup \texttt{LogRange}(10^{-4}, 10^{0}; 30)$
Shift penalty	$\lambda \in \{0\} \cup \bigcup_{i=0}^{2} \texttt{LinRange}(10^{i}, 10^{i+1}; 10) \cup \{\infty\}$

Table 6.7: **Portfolio-Optimization Performance.** Standard errors (± terms) reported in "Difference from SAA" row computed using common random numbers.

	SAA	Windowing	Smoothing	WPF with metric		
				L_1	L_2	L_∞
Risk-adjusted average testing cost	0.0707	0.0930	0.0896	0.0630	0.0772	0.0644
Difference from SAA (%)		31.6 ± 9.4	26.8 ± 8.0	-10.8 ± 5.8	9.2 ± 5.3	-8.8 ± 5.6

Figure 6.8 graphs average costs of the WPF method for different shift penalties, and Figure 6.9 graphs the probabilities assigned to the sequence of observations from Figure 6.7 by WPF when optimizing the portfolio.

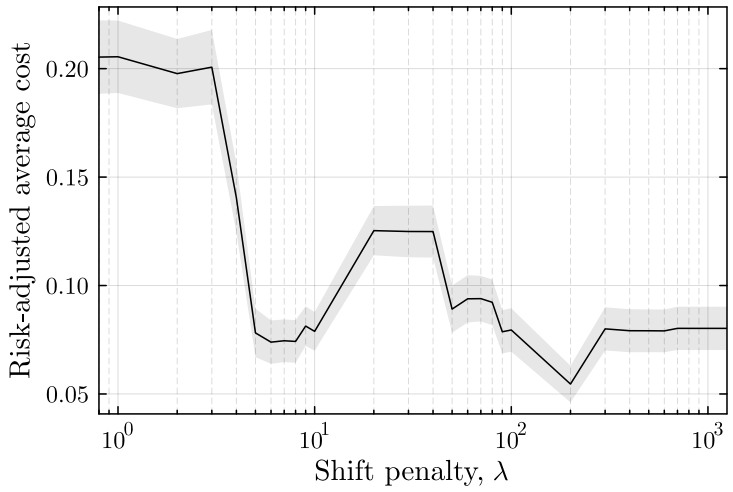

Figure 6.8: **Portfolio-Optimization Parameter Tuning for WPF with L_1 Metric.** Risk-adjusted average cost within the two-year parameter-tuning window. Band presents standard-error range.

In contrast to Figure 6.5, Figure 6.8 shows two decreases in costs for intermediate penalties.

This is evidence of different nonstationary processes fitting the data reasonably well, reflecting the multiple shocks observed at different times throughout Figure 6.7.

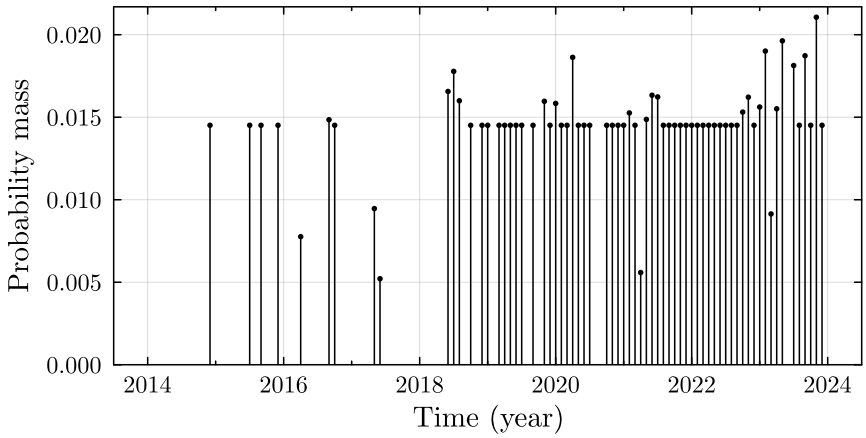

Figure 6.9: **Portfolio-Optimization Probabilities for WPF with L_1 Metric.** Shift penalty set at the minimum from Figure 6.8.

Figure 6.9 shows most probability assigned to observations after 2019. This is consistent with the two main shocks in Figure 6.7 occurring in 2020 and 2022, and the lower of the two local minima in Figure 6.8 being at a larger shift penalty.

6.4 Discussion

We have shown how the maximum-likelihood formulation with Wasserstein-distance regularization gives a problem that can be reformulated to have a simple network-flow structure. In three examples with nonstationary data, we demonstrate how effective this approach can be in practice. In particular, we show that the WPF method can be used to improve the forecasting of the dairy prices from Chapter 4, and this could thus be used to improve the real-world performance of the MPC scheme developed therein. We achieved this with the L_1 metric, which here performs better than the Euclidean metric. From a maximum-likelihood perspective, if the probability of a change in the distribution across multiple dimensions is simply the product of the probabilities of the change across each individual dimension, then after taking logarithms we get a framework in which the L_1 metric appears naturally. Even though in our two example data sets we do not expect the required dimension independence to hold, the results suggest that the L_1 metric may be a better choice than alternatives.

The flexibility of the WPF method with regard to different distance metrics is an advantage. We have already seen that the L_1 metric can be much more effective than the Euclidean metric when dealing with multidimensional data. Another option is to adjust an existing metric to

penalise any movement — thus we introduce a new parameter $\delta > 0$ and define the adjusted distance by

$$d'(\xi, \zeta) \coloneqq \begin{cases} d(\xi, \zeta) + \delta & \text{if } \xi \neq \zeta, \\ 0 & \text{if } \xi = \zeta. \end{cases}$$

As long as d is a lower-semicontinuous metric, it can be seen that d' is a lower-semicontinuous metric as well. This change introduces a sort of "stickiness", with more of the distribution being left unchanged from one period to the next. We found that this can produce marginal improvements in our numerical examples.

The WPF method results in an assignment of nonzero probabilities to a subset of the previous observations as an estimate of the final probability distribution. Depending on the value of the shift penalty, the number of observations included in this set may be quite small. The WPF method thus provides a suitable methodology for scenario reduction in the presence of nonstationarity. Existing approaches to scenario reduction focus on selecting a subset of data points that give a good representation of the overall distribution, where a Wasserstein distance can be used to measure the fit [98]. The need to make a small selection of scenarios from the complete historical record is driven by computational necessity in areas like energy planning [87].

The model we use has a single penalty parameter λ, corresponding to a situation in which the likelihood of changes in the underlying distribution is constant over time. So although we have nonstationarity, the characteristics of this nonstationarity do not vary over time. For example, our model could be applied in a case where the underlying distribution in each time period is chosen from a finite set of possible distributions and movement between these candidate distributions follows a Markov chain. The Wasserstein penalty would then be appropriate if the probability of moving from one state to another depends on the Wasserstein distance between the corresponding distributions. Then our assumption of a single value for λ corresponds to a fixed transition matrix. In our empirical examples we see variations in λ over the period of the time series, which suggests that *higher-order* nonstationarity — where the dynamics of the nonstationarity vary over time — may be important. We have not attempted a systematic exploration of this, but it seems that explicitly discarding older observations could be effective in these circumstances.

Finally, compared with the Wasserstein distributionally robust optimization model we developed in the previous chapter (the concentration-bound machinery of which led to *monotonic* observation weights), the WPF method has a distinct advantage in being able to adjust

weights *non*-monotonically and on a per-data outcome basis. But, the WPF method is *not* a distributionally robust approach; its weights can vary significantly with changes to the shift penalty parameter, and hence its performance is very sensitive to the tuning of this parameter. It would be interesting and worthwhile then to take a blended approach, where the weights are still non-monotonic and dependent on the data outcomes, but where they are affected by some notion of distributional robustness.

Chapter 7

Discussion

This thesis studies the theory and practice of data-driven decision making. Motivated by a real-world supply chain–management problem arising in the New Zealand dairy industry, we examine how to construct effective decision-making policies from limited and nonstationary data.

In Part I of the thesis, we study the performance of data-driven approximations of stochastic optimal control problems. Assuming that the historical data consist of independent and identically distributed samples from a stationary distribution, we establish general conditions under which sample-based approximations asymptotically converge to optimality. These conditions — based on the asymptotic integrability and semicontinuity of Bellman operators — are abstract but broadly applicable, and can be verified in practice using problem-specific features.

Beyond asymptotic convergence, we study the out-of-sample performance of different approximation schemes in the small-sample regime. In particular we compare stochastic dynamic programming and model predictive control, showing that model predictive control can be interpreted as solving a mean-constrained distributionally robust stochastic dynamic programming problem when its Bellman operator is concavity preserving. This plays a critical role in determining which approximation is suited out of sample, and we validate this insight through a revenue optimization problem with stochastic prices. However, the assumption of stationarity in the underlying price process is a limitation, motivating the second part of the thesis.

In Part II of the thesis, we study the estimation of probability distributions from data when the data-generating process is nonstationary. In this setting historical samples are no longer identically distributed, and may become less relevant over time. We thus develop a distributionally robust optimization model that centres Wasserstein ambiguity balls at weighted empirical distributions, thereby allowing for the time decay of information due to distributional drift.

This reveals a fundamental variance–drift trade-off, and in optimally balancing this trade-off, we derive weightings with a monotonic-polynomial structure.

To capture information contained in particular data outcomes, we further introduce a maximum likelihood formulation with a Wasserstein distance–based regularization term. The resulting estimation problem can be reformulated as a solvable finite-dimensional convex exponential-cone program. By appropriately tuning the regularization parameter to balance temporal smoothness against the structure of abrupt changes in the dynamics of the data, the approach is effective in many different applications which exhibit the complexities of data encountered in the wild.

Putting the understanding gained from Part I to work, in Chapter 4 we develop an approximation scheme suited to the characteristics of a supply chain problem from the New Zealand dairy industry. Noting that the problem leads to concavity-preserving Bellman operators, we address price uncertainty using a model predictive control approach built on a multiplicative vector-autoregressive forecasting model. In simulation this improves decision and predictive performance by 11 % and 40 %, respectively. We also note that a stochastic dual dynamic programming approach would be suited to address milk-supply uncertainty.

Putting the developments of Part II to work, in Chapter 6 we rise to the challenge of the price data facing the New Zealand dairy industry. We address nonstationarity using our Wasserstein-distance regularized maximum-likelihood approach, where data with similar dynamics to the present are selectively upweighted. In historical backtesting this improves forecasting performance by 8 %, and it therefore offers a practical enhancement to the forecasting-based model predictive control policies of Part I.

Taken together, the two parts of this thesis lay the foundations for data-driven stochastic control in data-scarce and nonstationary environments. While the methods developed herein deal with nonstationarity in a principled manner, they do so in a one-step look-ahead fashion, constructing an estimate of only the current (and prior) distributions based on historical data. Though this can be incorporated into a model predictive control or stochastic dynamic programming approximation through the use of a straightforward repeated point or distributional forecast, it neglects the obvious fact that, if the environment is changing over time, it is likely to continue to change over time into the future. An interesting direction for future research is then how to approach the estimation of a future *sequence* of nonstationary distributions, either as a sequence of point forecasts, or at the heart of some sequence of wider ambiguity sets.

In order to capture a sense of the momentum with which a nonstationary distribution is changing, and therefore how it is seemingly likely to continue changing in the future, it seems as though more of an (at least) semi-parametric structure would need to be imposed, since the space of probability distributions does not admit a vector-space structure. For instance, one could imagine modelling and extrapolating the momentum of the mean vector, while continuing with a Wasserstein-distance regularized maximum-likelihood approach to the estimation of the rest of the underlying distributions. How far to extrapolate such momentum into the future would then be a question of the efficacy of the resulting model.

Bibliography

[1] Ravindra K. Ahuja, Thomas L. Magnanti, & James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. 1st ed. Prentice Hall, 1993.

[2] Hadar Amrani & Eugene Khmelnitsky. "Estimation of quantiles of non-stationary demand distributions." *IISE Transactions* 49.4 (2017), pp. 381–394.

[3] Lin An, Andrew A. Li, Benjamin Moseley, & R. Ravi. "The nonstationary newsvendor with (and without) predictions." *Manufacturing & Service Operations Management* 27.3 (2025), pp. 881–896.

[4] Edward Anderson & Harrison Nguyen. "When can we improve on sample average approximation for stochastic optimization?" *Operations Research Letters* 48.5 (2020), pp. 566–572.

[5] Edward J. Anderson & Dominic S. T. Keehan. "Nonstationary distribution estimation via Wasserstein probability flows." Preprint. 2025.

[6] Edward J. Anderson & Andrew B. Philpott. "Improving sample average approximation using distributional robustness." *INFORMS Journal on Optimization* 4.1 (2022), pp. 90–124.

[7] Zvi Artstein & Roger J-B Wets. "Consistency of minimizers and the SLLN for stochastic programs." *Journal of Convex Analysis* 2.1-2 (1995), pp. 1–17.

[8] Hédy Attouch, Roberto Lucchetti, & Roger J-B Wets. "The topology of the ρ-Hausdorff distance." *Annali di Matematica Pura ed Applicata* CLX (1991), pp. 303–320.

[9] Hédy Attouch & Roger J-B Wets. "Isometries for the Legendre–Fenchel transform." *Transactions of the American Mathematical Society* 296.1 (1986), pp. 33–60.

[10] Hédy Attouch & Roger J-B Wets. "Quantitative stability of variational systems: I. The epigraphical distance." *Transactions of the American Mathematical Society* 328.2 (1991), pp. 695–729.

[11] Australia and New Zealand Banking Group Limited. *Commodity Price Index*. Website. Link: `anz.co.nz/about-us/economic-markets-research/commodity-price-index/`. 2024.

[12] Güzin Bayraksan & David K. Love. "Data-driven stochastic programming using ϕ-divergences." *The Operations Research Revolution*. TutORials in Operations Research. INFORMS, 2015, pp. 1–19.

[13] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[14] Dimitri P. Bertsekas. "Dynamic programming and suboptimal control: A survey from ADP to MPC." *European Journal of Control* 11.4 (2005), pp. 310–334.

[15] Dimitri P. Bertsekas. "Model predictive control and reinforcement learning: A unified framework based on dynamic programming." *IFAC-PapersOnLine* 58.18 (2024). 8th IFAC Conference on Nonlinear Model Predictive Control NMPC 2024, pp. 363–383.

[16] Dimitri P. Bertsekas. *Reinforcement Learning and Optimal Control*. Athena Scientific, 2019.

[17] Dimitri P. Bertsekas. *Rollout, Policy Iteration, and Distributed Reinforcement Learning*. Athena Scientific, 2021.

[18] Dimitri P. Bertsekas & Steven E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, 1996.

[19] Jeff Bezanson, Alan Edelman, Stefan Karpinski, & Viral B. Shah. "Julia: A fresh approach to numerical computing." *SIAM Review* 59.1 (2017), pp. 65–98.

[20] Patrick Billingsley. *Convergence of Probability Measures*. 2nd ed. Wiley, 1999.

[21] Patrick Billingsley. *Probability and Measure*. 3rd ed. Wiley, 1995.

[22] Jose Blanchet, Yang Kang, & Karthyek Murthy. "Robust Wasserstein profile inference and applications to machine learning." *Journal of Applied Probability* 56.3 (2019), pp. 830–857.

[23] Tim Bollerslev, Ray Y. Chou, & Kenneth F. Kroner. "ARCH modeling in finance: A review of the theory and empirical evidence." *Journal of Econometrics* 52.1-2 (1992), pp. 5–59.

[24] Stéphane Boucheron, Gábor Lugosi, & Pascal Massart. "Concentration inequalities using the entropy method." *The Annals of Probability* 31.3 (2003), pp. 1583–1614.

[25] Jacob Boudoukh, Matthew Richardson, & Robert Whitelaw. "The best of both worlds." *Risk* 11.5 (1998), pp. 64–67.

[26] Chris Chatfield, Anne B. Koehler, J. Keith Ord, & Ralph D. Snyder. "A new look at models for exponential smoothing." *Journal of the Royal Statistical Society: Series D (The Statistician)* 50.2 (2001), pp. 147–159.

[27] Boxiao Chen. "Data-driven inventory control with shifting demand." *Production and Operations Management* 30.5 (2021), pp. 1365–1385.

[28] Wang Chi Cheung, David Simchi-Levi, & Ruihao Zhu. "Nonstationary reinforcement learning: The blessing of (more) optimism." *Management Science* 69.10 (2023), pp. 5722–5739.

[29] Dairy Companies Association of New Zealand. *The New Zealand Dairy Industry*. Website. Link: `dcanz.com/the-new-zealand-dairy-industry/`. 2023.

[30] Herbert A. David & H. N. Nagaraja. *Order Statistics*. 3rd ed. Wiley, 2004.

[31] Russell Davidson & James G. MacKinnon. *Econometric Theory and Methods*. Oxford University Press, 2004.

[32] Russell Davidson & James G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, 1993.

[33] Erick Delage & Yinyu Ye. "Distributionally robust optimization under moment uncertainty with application to data-driven problems." *Operations Research* 58.3 (2010), pp. 595–612.

[34] Xialiang Dou & Mihai Anitescu. "Distributionally robust optimization with correlated data from vector autoregressive processes." *Operations Research Letters* 47.4 (2019), pp. 294–299.

[35] Anthony Downward, Oscar Dowson, & Reagan Baucke. "Stochastic dual dynamic programming with stagewise-dependent objective uncertainty." *Operations Research Letters* 48.1 (2020), pp. 33–39.

[36] Oscar Dowson. "Applying stochastic optimisation to the New Zealand dairy industry." Ph.D. thesis. University of Auckland, 2018.

[37] Oscar Dowson. "The policy graph decomposition of multistage stochastic programming problems." *Networks* 76.1 (2020), pp. 3–23.

[38] Oscar Dowson & Lea Kapelevich. "SDDP.jl: A Julia package for stochastic dual dynamic programming." *INFORMS Journal on Computing* 33.1 (2021), pp. 27–33.

[39] Dowson Farms. *Forecasting the 2024/25 Fonterra Milk Price*. Website. Link: `dairyanalytics.co.nz`. 2025.

[40] John C. Duchi, Peter W. Glynn, & Hongseok Namkoong. "Statistics of robust optimization: A generalized empirical likelihood approach." *Mathematics of Operations Research* 46.3 (2021), pp. 946–969.

[41] Iain Dunning, Joey Huchette, & Miles Lubin. "JuMP: A modeling language for mathematical optimization." *SIAM Review* 59.2 (2017), pp. 295–320.

[42] Jitka Dupačová. "On minimax solutions of stochastic linear programming problems." *Časopis pro pěstování matematiky* 91.4 (1966), pp. 423–430.

[43] Richard Durrett. *Probability: Theory and Examples*. 5th ed. Cambridge University Press, 2019.

[44] Martin Dyer & Leen Stougie. "Computational complexity of stochastic programming problems." *Mathematical Programming* 106 (2006), pp. 423–432.

[45] Víctor Elvira, Luca Martino, & Christian P. Robert. "Rethinking the effective sample size." *International Statistical Review* 90.3 (2022), pp. 525–550.

[46] Peyman Mohajerin Esfahani & Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations." *Mathematical Programming* 171 (2018), pp. 115–166.

[47] Eugene A. Feinberg, Pavlo O. Kasyanov, & Johannes O. Royset. "Epi-convergence of expectation functions under varying measures and integrands." *Journal of Convex Analysis* 30.3 (2023), pp. 917–936.

[48] Eugene A. Feinberg, Pavlo O. Kasyanov, & Nina V. Zadoianchuk. "Average cost Markov decision processes with weakly continuous transition probabilities." *Mathematics of Operations Research* 37.4 (2012), pp. 591–607.

[49] Eugene A. Feinberg, Pavlo O. Kasyanov, & Michael Z. Zgurovsky. "Partially observable total-cost Markov decision processes with weakly continuous transition probabilities." *Mathematics of Operations Research* 41.2 (2016), pp. 656–681.

[50] Nicolas Fournier. "Convergence of the empirical measure in expected Wasserstein distance: Non-asymptotic explicit bounds in $\mathbb{R}^d$." *ESAIM: PS* 27 (2023), pp. 749–775.

[51] Nicolas Fournier & Arnaud Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure." *Probability Theory and Related Fields* 162 (2015), pp. 707–738.

[52] Rui Gao. "Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality." *Operations Research* 71.6 (2023), pp. 2291–2306.

[53] Dongdong Ge, Qi Huangfu, Zizhuo Wang, Jian Wu, & Yinyu Ye. *Cardinal Optimizer (COPT) User Guide*. 2024.

[54] Pierre Girardeau, Vincent Leclere, & Andrew B. Philpott. "On the convergence of decomposition methods for multistage stochastic convex programs." *Mathematics of Operations Research* 40.1 (2014), pp. 130–145.

[55] Clark R. Givens & Rae M. Shortt. "A class of Wasserstein metrics for probability distributions." *Michigan Mathematical Journal* 31.2 (1984), pp. 231–240.

[56] Alison Goffin. *Farmers' Mental Health: A Review of the Literature*. Technical report. Accident Compensation Corporation, 2014.

[57] Jun-ya Gotoh, Michael Jong Kim, & Andrew E. B. Lim. "Technical note—A data-driven approach to beating SAA out of sample." *Operations Research* 73.2 (2023), pp. 829–841.

[58] Ziming Guan. *SDDP Production Planning Model and SDDP Hedge Contract Model*. Technical report. Fonterra, 2019.

[59] Ziming Guan. "Strategic inventory models for international dairy commodity markets." Ph.D. thesis. University of Auckland, 2008.

[60] Ziming Guan & Andrew B. Philpott. "A multistage stochastic programming model for the New Zealand dairy industry." *International Journal of Production Economics* 134.2 (2011), pp. 289–299.

[61] Yanru Guo, Ruiwei Jiang, & Siqian Shen. "Optimization with multi-sourced reference information and unknown trust: A distributionally robust approach." Preprint. 2025.

[62] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2025.

[63] Jie Hu, Zhi Chen, & Shuming Wang. “Budget-driven multiperiod hub location: A robust time-series approach.” *Operations Research* 73.2 (2024), pp. 613–631.

[64] Chengpiao Huang & Kaizheng Wang. “A stability principle for learning under nonstationarity.” *Operations Research* 73.6 (2025), pp. 3044–3064.

[65] Jeffrey Humpherys, Preston Redd, & Jeremy West. “A fresh look at the Kalman filter.” *SIAM Review* 54.4 (2012), pp. 801–823.

[66] Alexander D. Ioffe. “On lower semicontinuity of integral functionals. I.” *SIAM Journal on Control and Optimization* 15.4 (1977), pp. 521–538.

[67] Garud N. Iyengar. “Robust dynamic programming.” *Mathematics of Operations Research* 30.2 (2005), pp. 257–280.

[68] Caleb Ju & Guanghui Lan. “Dual dynamic programming for stochastic programs over an infinite horizon.” Preprint. 2025.

[69] Dominic S. T. Keehan, Edward J. Anderson, & Wolfram Wiesemann. “Don’t look back in anger: Wasserstein distributionally robust optimization with nonstationary data.” Preprint. 2025.

[70] Dominic S. T. Keehan, Andrew B. Philpott, & Edward J. Anderson. “On the out-of-sample performance of stochastic dynamic programming and model predictive control.” *INFORMS Journal on Optimization* 8.1 (2025), pp. 41–60.

[71] Dominic S. T. Keehan & Johannes O. Royset. “Epi-consistent approximation of stochastic dynamic programs.” *Journal of Convex Analysis* (2025). Forthcoming.

[72] N. Bora Keskin, Xu Min, & Jing-Sheng Jeannette Song. “The nonstationary newsvendor: Data-driven nonparametric learning.” Preprint. 2025.

[73] N. Bora Keskin & Assaf Zeevi. “Chasing demand: Learning and earning in a changing environment.” *Mathematics of Operations Research* 42.2 (2017), pp. 277–307.

[74] Alan J. King & Roger J-B Wets. “Epi-consistency of convex stochastic programs.” *Stochastics and Stochastic Reports* 34.1-2 (1991), pp. 83–92.

[75] Achim Klenke. *Probability Theory: A Comprehensive Course*. 3rd ed. Springer, 2020.

[76] Anton J. Kleywegt, Alexander Shapiro, & Tito Homem-de-Mello. “The sample average approximation method for stochastic discrete optimization.” *SIAM Journal on Optimization* 12.2 (2001), pp. 479–502.

[77] Augustine Kong. *A Note on Importance Sampling using Standardized Weights*. Technical report 348. Department of Statistics, University of Chicago, 1992.

[78] Lisa A. Korf. "Approximating infinite horizon stochastic optimal control in discrete time with constraints." *Annals of Operations Research* 142 (2006), pp. 165–186.

[79] Daniel Kuhn, Soroosh Shafiee, & Wolfram Wiesemann. "Distributionally robust optimization." *Acta Numerica* 34 (2025), pp. 579–804.

[80] Benoît Legat, Oscar Dowson, Joaquim Dias Garcia, & Miles Lubin. "MathOptInterface: A data structure for mathematical optimization problems." *INFORMS Journal on Computing* 34.2 (2021), pp. 672–689.

[81] Mengmeng Li, Tobias Sutter, & Daniel Kuhn. "Distributionally robust optimization with Markovian data." *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila & Tong Zhang. Vol. 139. PMLR, 2021, pp. 6493–6503.

[82] Bar Light. "The principle of optimality in dynamic programming: A pedagogical note." *Operations Research Letters* 57 (2024), p. 107164.

[83] Miles Lubin, Oscar Dowson, Joaquim Dias Garcia, Joey Huchette, Benoît Legat, & Juan Pablo Vielma. "JuMP 1.0: Recent improvements to a modeling language for mathematical optimization." *Mathematical Programming Computation* 15 (2023), pp. 581–589.

[84] Thomas Martin. "Stochastic optimization for the procurement of crude oil in refineries." Ph.D. thesis. École des Ponts ParisTech, 2021.

[85] Giacomo Nannicini, Emiliano Traversi, & Roberto Wolfler Calvo. "A Benders squared (B2) framework for infinite-horizon stochastic linear programs." *Mathematical Programming Computation* 13 (2020), pp. 645–681.

[86] François Pacaud, Pierre Carpentier, Jean-Philippe Chancelier, & Michel De Lara. "Optimization of a domestic microgrid equipped with solar panel and battery: Model predictive control and stochastic dual dynamic programming approaches." *Energy Systems* 15 (2024), pp. 115–139.

[87] SangWoo Park, Qingyu Xu, & Benjamin F. Hobbs. "Comparing scenario reduction methods for stochastic transmission planning." *IET Generation, Transmission & Distribution* 13.7 (2019), pp. 1005–1013.

[88] Teemu Pennanen. "Epi-convergent discretizations of multistage stochastic programs." *Mathematics of Operations Research* 30.1 (2005), pp. 245–256.

[89] Teemu Pennanen. "Epi-convergent discretizations of multistage stochastic programs via integration quadratures." *Mathematical Programming* 116 (2009), pp. 461–479.

[90] M. V. F. Pereira & L. M. V. G. Pinto. "Multi-stage stochastic optimization applied to energy planning." *Mathematical Programming* 52 (1991), pp. 359–375.

[91] Andrew B. Philpott & Ziming Guan. "On the convergence of stochastic dual dynamic programming and related methods." *Operations Research Letters* 36.4 (2008), pp. 450–455.

[92] Iosif Pinelis. "On the extreme points of moments sets." *Mathematical Methods of Operations Research* 83.3 (2016), pp. 325–349.

[93] Chi Seng Pun, Tianyu Wang, & Zhenzhen Yan. "Data-driven distributionally robust CVaR portfolio optimization under a regime-switching ambiguity set." *Manufacturing & Service Operations Management* 25.5 (2023), pp. 1779–1795.

[94] Ralph T. Rockafellar & Roger J-B Wets. *Variational Analysis*. Springer, 1998.

[95] Johannes O. Royset. "Approximations and solution estimates in optimization." *Mathematical Programming* 170 (2018), pp. 479–506.

[96] Johannes O. Royset & Roger J-B Wets. *An Optimization Primer*. Springer, 2021.

[97] Johannes O. Royset & Roger J-B Wets. "Variational analysis of constrained M-estimators." *The Annals of Statistics* 48.5 (2020), pp. 2759–2790.

[98] Napat Rujeerapaiboon, Kilian Schindler, Daniel Kuhn, & Wolfram Wiesemann. "Scenario reduction revisited: Fundamental limits and guarantees." *Mathematical Programming* 191 (2022), pp. 207–242.

[99] Yves Rychener, Adrian Esteban-Perez, Juan M. Morales, & Daniel Kuhn. "Wasserstein distributionally robust optimization with heterogeneous data sources." Preprint. 2024.

[100] Herbert E. Scarf. "A min-max solution of an inventory problem." *Studies in the Mathematical Theory of Inventory and Production*. Ed. by Kenneth J. Arrow, Samuel Karlin, & Herbert E. Scarf. Stanford University Press, 1958, pp. 201–209.

[101] George A. F. Seber & Alan J. Lee. *Linear Regression Analysis*. 2nd ed. Wiley, 2003.

[102] Aras Selvi, Eleonora Kreacic, Mohsen Ghassemi, Vamsi K. Potluru, Tucker Balch, & Manuela Veloso. "Distributionally and adversarially robust logistic regression via intersecting Wasserstein balls." *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*. Ed. by Silvia Chiappa & Sara Magliacane. Vol. 286. PMLR, 2025, pp. 3641–3674.

[103] Alexander Shapiro. "Inference of statistical bounds for multistage stochastic programming problems." *Mathematical Methods of Operational Research* 58 (2003), pp. 57–68.

[104] Alexander Shapiro. "On complexity of multistage stochastic programs." *Operations Research Letters* 34.1 (2006), pp. 1–8.

[105] Alexander Shapiro. "Rectangular sets of probability measures." *Operations Research* 64.2 (2016), pp. 528–541.

[106] Alexander Shapiro & Yi Cheng. "Central limit theorem and sample complexity of stationary stochastic programs." *Operations Research Letters* 49.5 (2021), pp. 676–681.

[107] Alexander Shapiro & Yi Cheng. "Dual bounds for periodical stochastic programs." *Operations Research* 71.1 (2023), pp. 120–128.

[108] Alexander Shapiro, Darinka Dentcheva, & Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.

[109] Alexander Shapiro & Lingquan Ding. "Periodical multistage stochastic programs." *SIAM Journal on Optimization* 30.3 (2020), pp. 2083–2102.

[110] Alexander Shapiro & Arkadi Nemirovski. "On complexity of stochastic programming problems." *Continuous Optimization: Current Trends and Modern Applications*. Ed. by Vaithilingam Jeyakumar & Alexander Rubinov. Springer, 2005, pp. 111–146.

[111] Jemma Simmonds. "Stochastic supply chain modelling in Julia." Honours thesis. University of Auckland, 2017.

[112] James E. Smith & Robert L. Winkler. "The optimizer's curse: Skepticism and postdecision surprise in decision analysis." *Management Science* 52.3 (2006), pp. 311–322.

[113] Nancy L. Stokey, Robert E. Lucas Jr., & Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.

[114] Tobias Sutter, Bart P. G. Van Parys, & Daniel Kuhn. "A Pareto dominance principle for data-driven optimization." *Operations Research* 72.5 (2024), pp. 1976–1999.

[115] Leonard J. Tashman. "Out-of-sample tests of forecasting accuracy: An analysis and review." *International Journal of Forecasting* 16.4 (2000), pp. 437–450.

[116] Bahar Taskesen, Man-Chung Yue, Jose Blanchet, Daniel Kuhn, & Viet Anh Nguyen. "Sequential domain adaptation by synthesizing distributionally robust experts." *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila & Tong Zhang. Vol. 139. PMLR, 2021, pp. 10162–10172.

[117] James T. Treharne & Charles R. Sox. "Adaptive inventory control for nonstationary demand and partial information." *Management Science* 48.5 (2002), pp. 607–624.

[118] Cédric Villani. *Optimal Transport: Old and New*. 1st ed. Springer, 2008.

[119] Minchao C. Yue, Daniel Kuhn, & Wolfram Wiesemann. "On linear optimization over Wasserstein balls." *Mathematical Programming* 195 (2022), pp. 1107–1122.