CrossMark

Incorporating statistical model error into the calculation of acceptability prices of contingent claims

Martin Glanzer[1] · Georg Ch. Pflug[1] · Alois Pichler[2]

Abstract

The determination of acceptability prices of contingent claims requires the choice of a stochastic model for the underlying asset price dynamics. Given this model, optimal bid and ask prices can be found by stochastic optimization. However, the model for the underlying asset price process is typically based on data and found by a statistical estimation procedure. We define a confidence set of possible estimated models by a nonparametric neighborhood of a baseline model. This neighborhood serves as ambiguity set for a multistage stochastic optimization problem under model uncertainty. We obtain distributionally robust solutions of the acceptability pricing problem and derive the dual problem formulation. Moreover, we prove a general large deviations result for the nested distance, which allows to relate the bid and ask prices under model ambiguity to the quality of the observed data.

✉ Georg Ch. Pflug
   georg.pflug@univie.ac.at

   Martin Glanzer
   martin.glanzer@univie.ac.at

   Alois Pichler
   alois.pichler@math.tu-chemnitz.de

1 Department of Statistics and OR (DSOR), University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

2 Faculty of Mathematics, Chemnitz University of Technology, Chemnitz, Germany

🙌 Springer

1 Introduction

The no-arbitrage paradigm is the cornerstone of mathematical finance. The fundamental work of Harrison, Kreps and Pliska [13–15,22] and Delbaen and Schachermayer [6], to mention some of the most important contributions, paved the way for a sound theory for the pricing of contingent claims. In a general market model, the exclusion of arbitrage opportunities leads to intervals of fair prices.

Typically, the resulting no-arbitrage price bounds are too wide to provide practically meaningful information.[1] In practice, market-makers wish to have a framework for controlling the acceptable risk when setting their spreads. Pioneering contributions to incorporate risk in the pricing procedure for contingent claims were made by Carr et al. [3] as well as Föllmer and Leukert [9,10], subsequent generalizations being made, e.g., by Nakano [24] or Rudloff [42]. The pricing framework of the present paper is in this spirit: by specifying acceptability functionals, an agent may control her shortfall risk in a rather intuitive manner. In particular, using the Average-Value-at-Risk ($\mathbb{AV@R}_\alpha$) will allow for a whole range of prices between the extreme cases of hedging with probability one (the traditional approach) and hedging w.r.t. expectation by varying the parameter α.

Nowadays, there is great awareness of the epistemic uncertainty inherent in setting up a stochastic model for a given problem. For single-stage and two-stage situations, there is a plethora of available literature on different approaches to account for model ambiguity (see the lists contained in [31, pp. 232–233] or [45, p. 2]). Recently, balls w.r.t. the Kantorovich–Wasserstein distance around an estimated model have gained a lot of popularity (e.g., [7,8,11,12,25,46]), while originally proposed by Pflug and Wozabal [34] in 2007. However, the literature on nonparametric ambiguity sets for multistage problems is still extremely sparse. Analui and Pflug [1] were the first to study balls w.r.t. the multistage generalization of the Kantorovich–Wasserstein distance, named nested distance,[2] for incorporating model uncertainty into multistage decision making. It is the aim of this article to further explore this rather uncharted territory. The classic mathematical finance problem of contingent claim pricing serves as a very well suited instance for doing so. In fact, while in the traditional pointwise hedging setup only the null sets of the stochastic model for the dynamics of the underlying asset price process influence the resulting price of a contingent claim, the full specification of the model affects the claim price when acceptability is introduced. Thus, model dependency is even stronger in the latter case, which is the topic of this paper.

Stochastic optimization offers a natural framework to deal with the problems of mathematical finance. Application of the fundamental work of Rockafellar and Wets [35–41] on conjugate duality and stochastic programming has led to a stream of literature on those topics. King [19] originally formulated the problem of contingent claim pricing as a stochastic program. Extensions of this approach have been made, amongst others, by King, Pennanen and their coauthors [18–21,26–28], Kallio and

---

[1] For example, the superreplication price for a plain vanilla call option in exponential Lévy models is given by the spot price of the underlying asset (see Cont and Tankov [4, Prop. 10.2]), which is a trivial upper bound for the call option price.

[2] The definition of the nested distance can be found in the "Appendix".

Ziemba [17] or Dahl [5]. The stochastic programming approach naturally allows for incorporating features and constraints of real-world markets and allows to efficiently obtain numerical results by applying the powerful toolkit of available algorithms for convex optimization problems.

The main contribution of this article is the link between statistical model error and the pricing of contingent claims, where the pricing methodology allows for a controlled hedging shortfall. The setup is inspired by practically very relevant aspects of decision making under both aleatoric and epistemic uncertainty. Given the stochastic model from which future evolutions are drawn, agents are willing to accept a certain degree of risk in their decisions. However, it may be dangerously misleading to neglect the fact that it is impossible to detect the true model without error. Thus, a distributionally robust framework, which takes the limitations of nonparametric statistical estimation into account, is required. In the statistical terminology, balls w.r.t. the nested distance may be seen as confidence regions: by considering all models whose nested distance to the estimated baseline model does not exceed some threshold, it is ensured that the true model is covered with a certain probability and hence the decision is robust w.r.t. the statistical model estimation error. In particular, we prove a large deviations theorem for the nested distance, based on which we show that a scenario tree can be constructed out of data such that it converges (in terms of the nested distance) to the true model in probability at an exponential rate. Thus, distributionally robust claim prices w.r.t. nested distance balls as ambiguity sets include a hedge under the true model with arbitrary high probability, depending on the available data. In other words, we provide a framework that allows for setting up bid and ask prices for a contingent claim which result from finding hedging strategies with truly calculated risks, since the important factor of model uncertainty is not neglected.

This paper is organized as follows. In Sect. 2 we introduce our framework for acceptability pricing, i.e., we replace the traditional almost sure super-/subreplication requirement by the weaker constraint of an acceptable hedge. The acceptability condition is formulated w.r.t. one given probability model. This lowers the ask price and increases the bid price such that the bid–ask spread may be tightened or even closed. Section 3 contains the main results of this article. We weaken the assumption of one single probability model assuming that a collection of models is plausible. In particular, we define the distributionally robust acceptability pricing problem and derive the dual problem formulation under rather general assumptions on the ambiguity set. The effect of the introduction of acceptability and ambiguity into the classical pricing methodology is nicely mirrored by the dual formulations. Moreover, we give a strong statistical motivation for using nested distance balls as ambiguity sets by proving a large deviations theorem for the nested distance. Section 4 contains illustrative examples to visualize the effect of acceptability and model ambiguity on contingent claim prices. In Sect. 5 we discuss the algorithmic solution of the $\mathbb{AV@R}$-acceptability pricing problem w.r.t. nested distance balls as ambiguity sets. In particular, we exploit the duality results of Sect. 3 and the special stagewise structure of the nested distance by a sequential linear programming algorithm which yields approximate solutions to the originally semi-infinite non-convex problem. In this way, we overcome the current state-of-the-art computational methods for multistage stochastic optimization prob-

lems under non-parametric model ambiguity. Finally, we summarize our results in Sect. 6.

2 Acceptability pricing

2.1 Acceptability functionals

The terminology introduced in this section follows the book of Pflug and Römisch [33]. A detailed discussion of acceptability functionals and their properties can be found therein. Intuitively speaking, an acceptability functional $\mathcal{A}$ maps a stochastic position $Y \in L_p(\Omega), 1 < p < \infty$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, to the real numbers extended by $-\infty$ in such a way that higher values of the position correspond to higher values of the functional, i.e., a 'higher degree of acceptance'. In particular, the defining properties of an acceptability functional are *translation equivariance*,[3] *concavity*, *monotonicity*,[4] and *positive homogeneity*. We assume all acceptability functionals to be *version independent*,[5] i.e., $\mathcal{A}(Y)$ depends only on the distribution of the random variable Y.

The following proposition is well-known. It follows directly from the Fenchel–Moreau–Rockafellar Theorem (see [35, Th. 5] and [33, Th. 2.31]).

Proposition 1 *An acceptability functional $\mathcal{A}$ which fulfills the above conditions has a dual representation of the form*

$$\mathcal{A}(Y) = \inf \{\mathbb{E}[YZ] : Z \in \mathcal{Z}\},$$

where $\mathcal{Z}$ is a closed convex subset of $L^q(\Omega)$, with $1/p + 1/q = 1$. We call $\mathcal{Z}$ the superdifferential of $\mathcal{A}$. Monotonicity and translation equivariance imply that all $Z \in \mathcal{Z}$ are nonnegative densities.

Assumption A1 There exists some constant $K_1 \in \mathbb{R}$ such that for all $Z \in \mathcal{Z}$ it holds $\|Z\|_q \leq K_1$.

This assumption implies that $\mathcal{A}$ is Lipschitz on L_p:

$$|\mathcal{A}(Y_1) - \mathcal{A}(Y_2)| \leq K_1 \|Y_1 - Y_2\|_p. \tag{1}$$

A good example for such an acceptability functional is the Average Value-at-Risk, $\mathbb{AV@R}_\alpha$, whose superdifferential is given by

$$\mathcal{Z} = \{Z \in L_1(\Omega) : 0 \leq Z \leq 1/\alpha \text{ and } \mathbb{E}(Z) = 1\}.$$

---

[3] $\mathcal{A}(Y + c) = \mathcal{A}(Y) + c$ for any $c \in \mathbb{R}$.

[4] $X \leq Y$ a.s. $\implies \mathcal{A}(X) \leq \mathcal{A}(Y)$.

[5] For version independent acceptability functionals, upper semi-continuity follows from concavity (see Jouini, Schachermayer and Touzi [16]).

The extreme cases are represented by the essential infimum ($\mathbb{AV@R}_0(Y) := \lim_{\alpha \downarrow 0} \mathbb{AV@R}_\alpha(Y) = \mathrm{essinf}(Y)$[6]) and the expectation ($\alpha = 1$). Its superdifferentials are given by the set of all probability densities and just the function identically 1, respectively.

Other common names for the $\mathbb{AV@R}$ are Conditional-Value-at-Risk, Tail-Value-at-Risk, or Expected Shortfall. The subtleties between these terminologies are, e.g., addressed in Sarykalin et al. [43]. All our computational studies in Sect. 4 and Sect. 5 will be based on some $\mathbb{AV@R}_\alpha$, while our theoretical results are general.

2.2 Acceptable replications

Let us now introduce the notion of acceptability in the pricing procedure for contingent claims.

As usual in mathematical finance, we consider a market model as a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where the filtration is given by the increasing sequence of sigma-algebras $\mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_T)$ with $\mathcal{F}_0 = \{\emptyset, \Omega\}$. The liquidly traded basic asset prices are given by a discrete-time $\mathbb{R}^m_+$-valued stochastic process $S = (S_0, \ldots, S_T)$, where $S_t = (S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(m)})$. We assume the filtration to be generated by the asset price process.

One asset, denoted by $S^{(1)}$, serves as numéraire (a risk-less bond, say). We assume w.l.o.g. that $S_t^{(1)} = 1$ a.s. If not, we may replace $(S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(m)})$ by $(1, S_t^{(2)}/S_t^{(1)}, \ldots, S_t^{(m)}/S_t^{(1)})$.

A contingent claim C consists of an $\mathcal{F}$-adapted series of cash flows $C = (C_1, \ldots, C_T)$ measured in units of the numéraire. The fact that the payoff C_t is contingent on the respective state of the market up to time t is reflected by the condition that C is adapted to the filtration $\mathcal{F}$, for which we write $C \lhd \mathcal{F}$. A trading strategy $x = (x_0, \ldots, x_{T-1})$ is an $\mathcal{F}$-adapted $\mathbb{R}^m$-valued process with $x \lhd \mathcal{F}$.

To be more precise, let

$$\mathcal{L}_p^m := \mathbb{R}^m \times L_p^m(\Omega, \mathcal{F}_1) \times \cdots \times L_p^m(\Omega, \mathcal{F}_T),$$
$$\mathcal{L}_\infty^m := \mathbb{R}^m \times L_\infty^m(\Omega, \mathcal{F}_1) \times \cdots \times L_\infty^m(\Omega, \mathcal{F}_{T-1}),$$

and

$$\mathcal{L}_q^1 := L_q(\Omega, \mathcal{F}_1) \times \cdots \times L_q(\Omega, \mathcal{F}_T).$$

We assume that $S \in \mathcal{L}_p^m$, $x \in \mathcal{L}_\infty^m$ and $C \in \mathcal{L}_p^1$. The norm in L_p^m is given by

$$\|Y\|_p = \sum_{i=1}^m \|Y^{(i)}\|_p,$$

and similarly for L_∞^m. Notice that x_0 and S_0 are deterministic vectors.

---

[6] Strictly speaking, Assumption A1 is not respected by $\mathbb{AV@R}_0$. However, all our results on $\mathbb{AV@R}$–acceptability pricing will hold true also for $\mathbb{AV@R}_0$. In fact, this is the special case which is well treated in the literature.

Assumption A2 We assume that all claims are Lipschitz-continuous functions of the underlying asset price process S.

Definition 1 Consider a contingent claim C and fix acceptability functionals $\mathcal{A}_t$, for all $t = 1, \ldots, T$. We assume that all functionals $\mathcal{A}$ have a representation given by Proposition 1. Then the acceptable prices are given by the optimal values of the following stochastic optimization programs:

(i) the **acceptable ask price** of C is defined as

$$
\text{(P)} \quad \left\| \begin{aligned}
\pi_a(\mathcal{A}_1, \ldots, \mathcal{A}_T) &= \min_x \; x_0^\top S_0 \\
\text{s.t. } & \mathcal{A}_t(x_{t-1}^\top S_t - x_t^\top S_t - C_t) \geq 0 \qquad\qquad (2a) \\
& \mathcal{A}_T(x_{T-1}^\top S_T - C_T) \geq 0, \qquad\qquad\qquad (2b)
\end{aligned} \right.
$$

(ii) the **acceptable bid price** of C is defined as

$$
\text{(P')} \quad \left\| \begin{aligned}
\pi_b(\mathcal{A}_1, \ldots, \mathcal{A}_T) &= \max_x \; x_0^\top S_0 \\
\text{s.t. } & \mathcal{A}_t(x_t^\top S_t - x_{t-1}^\top S_t + C_t) \geq 0 \qquad\qquad (3a) \\
& \mathcal{A}_T(-x_{T-1}^\top S_T + C_T) \geq 0, \qquad\qquad\quad (3b)
\end{aligned} \right.
$$

where the optimization runs over all trading strategies $x \in \mathcal{L}_\infty^m$ for the liquidly traded assets. The constraints in (2a) and (3a) are formulated for all $t = 1, \ldots, T - 1$.

To interpret Definition 1, the acceptable ask price is given by the minimal initial capital required to acceptably superhedge the cash-flows C_t, which have to be paid out by the seller. On the other hand, the acceptable bid price corresponds to the maximal amount of money that can initially be borrowed from the market to buy the claim, such that by receiving the payments C_t and always rebalancing one's portfolio in an acceptable way, one ends up with an acceptable position at maturity.

In what follows we will mainly consider the ask price problem (P) and its variants. The bid price problem (P') is its mirror image and all assertions and proofs for the problem (P) can be rewritten literally for problem (P').

Let (P$^\beta$) for $\beta = (\beta_1, \ldots, \beta_T)$ be the problem (P), where the conditions (2a) and (2b) are replaced by $\mathcal{A}_t(\cdot) \geq \beta_t$.

Assumption A3 The optima are attained and all solutions x to the problems (P$^\beta$), for β in a neighborhood of 0, are uniformly bounded, i.e., $\exists K_2 \in \mathbb{R}$ s.t. $\forall x : \|x\|_\infty \leq K_2$.

We show the following auxiliary result for the problems (P$^\beta$).

Lemma 1 *Let v^β be the optimal value of* (P$^\beta$) *and v^* be the optimal value of* (P). *Then, in a neighborhood of 0,*

$$
|v^\beta - v^*| \leq 2\bar{\beta} \cdot \|S_0\|_1 \qquad\qquad (4)
$$

where $\bar{\beta} = \sum_t |\beta_t|$.

Proof If v^β is the optimal value of (P^β), then by inclusion of the feasible sets

$$v^{-|\beta|} \le v^* \le v^{|\beta|},$$
$$v^{-|\beta|} \le v^\beta \le v^{|\beta|}.$$

We have to bound $v^{|\beta|} - v^{-|\beta|}$. Let x_t^* be the solution of $(P^{-|\beta|})$. x_t^* is not necessarily feasible for $(P^{|\beta|})$. We modify x_t^* in order to get feasibility for $(P^{|\beta|})$. Let $a_t, t = 1, \ldots, T-1$, be the vector with identical components $2 \sum_{s=t+1}^{T} |\beta_s|$ and let $x_t = x_t^* + a_t$. Then

$$\mathbb{E}[(x_{t-1} - x_t)^\top S_t Z_t] - \mathbb{E}[(x_{t-1}^* - x_t^*)^\top S_t Z_t]$$
$$= \mathbb{E}[(a_{t-1} - a_t)^\top S_t Z_t] = 2|\beta_t| \sum_{i=1}^{m} \mathbb{E}\left[S_t^{(i)} Z_t\right]$$
$$\ge 2|\beta_t| \cdot \left(\inf \sum_{i=1}^{m} S_t^{(i)}\right) \cdot \mathbb{E}[Z_t] \ge 2|\beta_t|$$

since $\sum_i S_t^{(i)} \ge S_t^{(1)} = 1$ and $\mathbb{E}[Z_t] = 1$. By $\mathbb{E}[(x_{t-1}^* - x_t^*)^\top S_t Z_t] \ge -|\beta_t|$, one gets that $\mathbb{E}[(x_{t-1} - x_t)^\top S_t Z_t] \ge |\beta_t|$, i.e., x_t is feasible for $(P^{|\beta|})$. Notice that a_0 has all components equal to $\sum_t |\beta_t| = \bar{\beta}$. Now

$$0 \le v^{|\beta|} - v^{-|\beta|} \le x_0^\top S_0 - x_0^{*\top} S_0 = a_0^\top S_0 = 2\bar{\beta} \sum_i S_0^{(i)} = 2\bar{\beta} \cdot \|S_0\|_1,$$

which concludes the proof. □

Notice that the primal program (P) is semi-infinite, if the constraints are written in the extensive form

$$\mathbb{E}\left[\left((x_{t-1} - x_t)^\top S_t - C_t\right) Z_t\right] \ge 0 \quad \text{for all } Z_t \in \mathcal{Z}_t,$$

where $Z = (Z_1, \ldots, Z_T) \in \mathcal{L}_q^1$.

Lemma 2 below demonstrates the validity of an approximation with only finitely many supergradients.

Since the L_p spaces are separable, there exist sequences $(Z_{t,1}, Z_{t,2}, \ldots)$ that are dense in $\mathcal{Z}_t$, for each t. Let

$$\mathcal{A}_{t,n}(Y) = \min\{\mathbb{E}[Y \cdot Z_{t,i}] : 1 \le i \le n\}.$$

Since $Z \mapsto \mathbb{E}[YZ]$ is continuous in L_p, for every Y in $L_p(\Omega, \mathcal{F}_t)$ it holds that

$$\mathcal{A}_{t,n}(Y) \downarrow \mathcal{A}_t(Y),$$

as $n \to \infty$.

Lemma 2 *Let v^* be the optimal value of the basic problem* (P) *and let v_n^* be the optimal value of the similar optimization problem* (P_n), *where $\mathcal{A}_t$ are replaced by $\mathcal{A}_{t,n}$. Then*

$$v_n^* \uparrow v^*.$$

Proof Suppose the contrary, that is $\sup_n v_n^* \le v^* - 3\eta < v^*$ for some $\eta > 0$. Introduce the notation

$$Y_t(x) = \begin{cases} (x_{t-1} - x_t)^\top S_t - C_t & \text{for } 1 \le t < T \\ x_{T-1}^\top S_T - C_T & \text{for } t = T. \end{cases}$$

By Assumption A1 and since $x \in \mathcal{L}_\infty^m$, it holds that $x \mapsto \mathcal{A}_t(Y_t(x))$ and $x \mapsto x_0^\top S_0$ are Lipschitz. Choose $0 < \delta = \eta \, [2\|S_0\|_1 K_1 (K_2 + K_3 + 1)]^{-1}$ with $K_3 \ge \|S_t\|_p$ for all t. Let x_t^* be the solution of (P). We may find finite sub-sigma-algebras $\tilde{\mathcal{F}}_t \subseteq \mathcal{F}_t$ such that with

$$\tilde{S}_t = \mathbb{E}[S_t | \tilde{\mathcal{F}}_t] \quad \text{(componentwise)},$$
$$\tilde{C}_t = \mathbb{E}[C_t | \tilde{\mathcal{F}}_t],$$
$$\tilde{x}_t^* = \mathbb{E}[x_t^* | \tilde{\mathcal{F}}_t] \quad \text{(componentwise)},$$

we have that

$$\|S_t - \tilde{S}_t\|_p \le \delta,$$
$$\|C_t - \tilde{C}_t\|_p \le \delta,$$
$$\|x_t^* - \tilde{x}_t^*\|_\infty \le \delta.$$

Denote by ($\tilde{\text{P}}$) the variant of the problem (P), where the processes (S_t) and (C_t) are replaced by $(\tilde{S}_t)$ and $(\tilde{C}_t)$. Similarly as before introduce the notation

$$\tilde{Y}_t(x) = \begin{cases} (x_{t-1} - x_t)^\top \tilde{S}_t - \tilde{C}_t & \text{for } 1 \le t < T \\ x_{T-1}^\top \tilde{S}_T - \tilde{C}_T & \text{for } t = T. \end{cases}$$

Notice that

$$
\begin{aligned}
&|\mathcal{A}_t(\tilde{Y}_t(\tilde{x}_t^*)) - \mathcal{A}_t(Y_t(x_t^*))| \\
&\le K_1 \|\tilde{Y}_t(\tilde{x}_t^*) - Y_t(x_t^*)\|_p \\
&\le K_1 \left[\|\tilde{x}_t^* - x_t^*\|_\infty \|\tilde{S}_t\|_p + \|x_t^*\|_\infty \|\tilde{S}_t - S_t\|_p + \|\tilde{C}_t - C_t\|_p \right] \\
&\le K_1 [\delta K_3 + \delta K_2 + \delta] = \eta \, [2\|S_0\|_1]^{-1}.
\end{aligned}
$$

By Lemma 1 we may conclude that

$$v^* \le \tilde{v}^* + \eta, \tag{5}$$

where $\tilde{v}^*$ is the optimal value of $(\tilde{P})$. Let $(\tilde{P}_n)$ be the variant of problem $(\tilde{P})$, where all $\mathcal{A}_t$ are replaced by $\mathcal{A}_{t,n}$. The optimal value of $(\tilde{P}_n)$ is denoted by $\tilde{v}_n^*$. In this finite situation we may show that $\tilde{v}_n^* \uparrow \tilde{v}^*$. Obviously, $\tilde{v}_n^*$ is a monotonically increasing sequence with $\tilde{v}_n^* \leq \tilde{v}^*$.

It remains to demonstrate that $\lim_n \tilde{v}_n^*$ cannot be smaller than $\tilde{v}^*$. For this, let $\tilde{x}^{n*}$ be a solution of $(\tilde{P}_n)$. Because of the finiteness of the filtration $\tilde{\mathcal{F}}$, the solutions of $(\tilde{P}_n)$ as well as of $\tilde{P}$ are just bounded vectors in some high-, but finite dimensional $\mathbb{R}^N$ and are all bounded by K_2. Let $\tilde{x}^{**}$ be an accumulation point of $(\tilde{x}^{n*})$, i.e., we have for some subsequence that $\tilde{x}^{n_i*} \to \tilde{x}^{**}$. We show that $\tilde{x}^{**}$ satisfies the constraints of $(\tilde{P})$.

Suppose the contrary. Then there is a t such that $\mathcal{A}_t(\tilde{Y}_t(\tilde{x}^{**})) < 0$. This implies that there is a $Z_{t,m} \in \{Z_{t,1}, Z_{t,2}, \ldots\}$ such that $\mathbb{E}[\tilde{Y}_t(\tilde{x}^{**}) \cdot Z_{t,m}] < 0$. However, for $n \geq m$, by construction $\mathbb{E}[\tilde{Y}_t(\tilde{x}^{n*}) \cdot Z_{t,m}] \geq 0$ and since $\tilde{x}^{n*} \to \tilde{x}^{**}$ componentwise, then also $\mathbb{E}[\tilde{Y}_t(\tilde{x}^{**}) \cdot Z_{t,m}] \geq 0$. Since the objective function is continuous in $\tilde{x}$ this implies that $\lim_i \tilde{v}_{n_i}^* = \tilde{v}^*$ and, by monotonicity, $\lim_n \tilde{v}_n^* = \tilde{v}^*$. We have therefore shown that we can find an index n such that

$$\tilde{v}^* < \tilde{v}_n^* + \eta. \tag{6}$$

Let x^{n*} be the solution of (P_n) and let $\hat{x}^{n*} = \mathbb{E}[x^{n*}|\tilde{\mathcal{F}}_t]$. Analogously as before, one may prove that $|\mathcal{A}_t(\tilde{Y}_t(\hat{x}^{n*}))| \leq \eta [2\|S_0\|_1]^{-1}$ and hence, by Lemma 1,

$$\tilde{v}_n^* \leq v_n^* + \eta. \tag{7}$$

Putting (5), (6) and (7) together one sees that

$$v^* \leq v_n^* + 3\eta,$$

which contradicts the assumption that $v_n^* < v^* - 3\eta$. $\qquad\square$

We now turn to the duals of the problems (P) and (P′), called (D) and (D′), respectively. It turns out that also in our general acceptability case a martingale property appears in the dual as it is known for the case of a.s. super-/subreplication.

Theorem 1 *For all $t = 1, \ldots, T$, let $\mathcal{A}_t$ be acceptability functionals with corresponding superdifferentials $\mathcal{Z}_t$. Then, the acceptable ask price is given by*

$$\text{(D)} \left\| \begin{array}{ll} \pi_a(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \displaystyle\sup_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}}\left[\sum_{t=1}^{T} C_t\right] & \\[12pt] \text{s.t. } \mathbb{E}^{\mathbb{Q}}[S_{t+1}|\mathcal{F}_t] = S_t & \forall t = 0, \ldots, T-1 \quad \text{(8a)} \\[12pt] \left.\dfrac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} \in \mathcal{Z}_t & \forall t = 1, \ldots, T, \quad \text{(8b)} \end{array} \right.$$

and the acceptable bid price is given by

$$
\text{(D}')\quad
\begin{Vmatrix}
\pi_b(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \inf_{\mathbb{Q}} \ \mathbb{E}^{\mathbb{Q}}\left[\sum_{t=1}^{T} C_t\right] \\[2mm]
\text{s.t. } \mathbb{E}^{\mathbb{Q}}[S_{t+1}|\mathcal{F}_t] = S_t \qquad \forall t = 0, \ldots, T-1 \\[2mm]
\left.\dfrac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} \in \mathcal{Z}_t \qquad \forall t = 1, \ldots, T.
\end{Vmatrix}
$$

(9a)

(9b)

Proof The acceptable ask/bid price corresponds to a special case of the distributionally robust acceptable ask/bid price introduced in Definition 2 below, namely when the ambiguity set reduces to a singleton. Hence, the validity of Theorem 1 follows directly from the proof of Theorem 2. □

Remark 1 (Interpretation of the dual formulations) The objective of the dual formulations (D) and (D′) is to maximize (minimize, resp.) the expected value of the payoffs resulting from the claim w.r.t. some feasible measure $\mathbb{Q}$. The constraints (8a) and (9a) require $\mathbb{Q}$ to be such that the underlying asset price process is a martingale w.r.t. $\mathbb{Q}$. This is well known from the traditional approach of pointwise super-/subreplication. The acceptability criterion enters the dual problems in terms of the constraints (8b) and (9b), which reduce the feasible sets by a stronger condition than the two probability measures just having the same null sets. Making the feasible sets smaller obviously lowers the ask price and increases the bid price and thus gives a tighter bid–ask spread.

Proposition 2 *For fixed acceptability functionals $\mathcal{A}_1, \ldots, \mathcal{A}_T$, consider the acceptable ask price $\pi^a(\mathbb{P})$ as a function of the underlying model $\mathbb{P}$. This function is Lipschitz.*

Proof The assertion follows from Theorem 5 in the "Appendix", considering the Lipschitz property of claims (Assumption A2) and the problem formulation resulting from Theorem 1. □

3 Model ambiguity and distributional robustness

Traditional stochastic programs are based on a given and fixed probability model for the uncertainties. However, already since the pioneering paper of Scarf [44] in the 1950s, it was felt that the fact that these models are based on observed data as well as the statistical error should be taken into account when making decisions. Ambiguity sets are typically either a finite collection of models or a neighborhood of a given baseline model. In what follows we study the latter case and, in particular, we use the nested distance to construct parameter-free ambiguity sets.

3.1 Acceptability pricing under model ambiguity

In Sect. 2.2 we defined the bid/ask price of a contingent claim as the maximal/minimal amount of capital needed in order to sub-/superhedge its payoff(s) w.r.t. an acceptability criterion. However, the result computed with this approach heavily depends

on the particular choice of the probability model. This section weakens the strong dependency on the model. More specifically, acceptable bid and ask prices shall be based on an acceptability criterion that is robust w.r.t. all models contained in a certain ambiguity set.

Definition 2 Consider a contingent claim C. Then, for acceptability functionals $\mathcal{A}_t$, $t = 1, \ldots, T$, and an ambiguity set $\mathcal{P}_\varepsilon$ of probability models,

(i) the **distributionally robust acceptable ask price** of C is defined as

$$
\text{(PP)} \quad \left\|
\begin{aligned}
&\pi_a^{\mathcal{P}_\varepsilon}(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \min_x \; x_0^\top S_0 \\
&\text{s.t. } \mathcal{A}_t^{\mathbb{P}}(x_{t-1}^\top S_t - x_t^\top S_t - C_t) \geq 0 \;\; \forall \mathbb{P} \in \mathcal{P}_\varepsilon \quad (10a) \\
&\qquad \mathcal{A}_T^{\mathbb{P}}(x_{T-1}^\top S_T - C_T) \geq 0 \;\; \forall \mathbb{P} \in \mathcal{P}_\varepsilon, \quad (10b)
\end{aligned}
\right.
$$

(ii) the **distributionally robust acceptable bid price** is defined as

$$
\text{(PP')} \quad \left\|
\begin{aligned}
&\pi_b^{\mathcal{P}_\varepsilon}(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \max_x \; x_0^\top S_0 \\
&\text{s.t. } \mathcal{A}_t^{\mathbb{P}}(x_t^\top S_t - x_{t-1}^\top S_t + C_t) \geq 0 \;\; \forall \mathbb{P} \in \mathcal{P}_\varepsilon \quad (11a) \\
&\qquad \mathcal{A}_T^{\mathbb{P}}(-x_{T-1}^\top S_T + C_T) \geq 0 \;\; \forall \mathbb{P} \in \mathcal{P}_\varepsilon, \quad (11b)
\end{aligned}
\right.
$$

where the optimization runs over all trading strategies $x \in \mathcal{L}_\infty^m$ for the liquidly traded assets. The constraints in (10a) and (11a) are formulated for all $t = 1, \ldots, T - 1$ and $\mathcal{A}_t^{\mathbb{P}}$ denotes the value of the acceptability functional when the underlying probability model is given by $\mathbb{P}$.

Theorem 2 *Let $\mathcal{P}_\varepsilon$ be a convex set of probability models, which is spanned by a sequence of models $(\mathbb{P}_1, \mathbb{P}_2, \ldots)$. Moreover, let $\mathcal{P}_\varepsilon$ be dominated by some model $\mathbb{P}_0$ and assume all densities w.r.t. $\mathbb{P}_0$ to be bounded. For $t = 1, \ldots, T$, let $\mathcal{A}_t$ be acceptability functionals with corresponding superdifferentials $\mathcal{Z}_{\mathcal{A}_t}$. Then, the distributionally robust acceptable ask price is given by*

$$
\text{(DD)} \quad \left\|
\begin{aligned}
&\pi_a^{\mathcal{P}_\varepsilon}(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \sup_{\mathbb{Q}} \; \mathbb{E}^{\mathbb{Q}}\left[\sum_{t=1}^T C_t \right] \\
&\text{s.t. } \mathbb{E}^{\mathbb{Q}}[S_{t+1} | \mathcal{F}_t] = S_t \qquad \forall t < T \quad (12a) \\
&\qquad \forall t \; \exists \mathbb{P} \in \mathcal{P}_\varepsilon : \left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} \in \mathcal{Z}_{\mathcal{A}_t^{\mathbb{P}}}, \quad (12b)
\end{aligned}
\right.
$$

and the distributionally robust acceptable bid price is given by

$$
\text{(DD')} \quad \left\|
\begin{aligned}
&\pi_b^{\mathcal{P}_\varepsilon}(\mathcal{A}_1, \ldots, \mathcal{A}_T) = \inf_{\mathbb{Q}} \; \mathbb{E}^{\mathbb{Q}}\left[\sum_{t=1}^T C_t \right] \\
&\text{s.t. } \mathbb{E}^{\mathbb{Q}}[S_{t+1} | \mathcal{F}_t] = S_t \qquad \forall t < T \quad (13a) \\
&\qquad \forall t \; \exists \mathbb{P} \in \mathcal{P}_\varepsilon : \left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} \in \mathcal{Z}_{\mathcal{A}_t^{\mathbb{P}}}. \quad (13b)
\end{aligned}
\right.
$$

Proof Define

$$\mathfrak{D}_t := \left\{ Z_t f_t \colon \exists\, \mathbb{P} \in \mathcal{P}_\varepsilon \text{ s.t. } Z_t \in \mathcal{Z}_{\mathcal{A}_t^\mathbb{P}},\ \left.\frac{d\mathbb{P}}{d\mathbb{P}_0}\right|_{\mathcal{F}_t} = f_t \right\}.$$

Then, the constraints in (PP′) can be written in the form

$$\mathbb{E}^{\mathbb{P}_0}[(x_{t-1} - x_t)^\top S_t - C_t)\partial_t] \geq 0 \quad \forall \partial_t \in \mathfrak{D}_t.$$

Since all densities f_t are bounded by assumption,[7] Lemma 2 holds true if we replace $Z_t \in \mathcal{Z}_t$ by $\partial_t \in \mathfrak{D}_t$. It can easily be seen that for each t there are sequences $(\partial_{t,1}, \partial_{t,2}, \ldots)$ which are dense in $\mathfrak{D}_t$. Let us define

$$\mathfrak{D}_t^n := \left\{ \sum_{i=1}^{n_1} \sum_{j=1}^{n_2^i} \lambda_{i,j} Z_t^{j,i} f_t^i \colon \sum_{i=1}^{n_1} \sum_{j=1}^{n_2^i} \lambda_{i,j} = 1, \left| \left\{ (i,j) : 1 \leq i \leq n_1, 1 \leq j \leq n_2^i \right\} \right| = n \right\}.$$

Then, it holds that $\mathfrak{D}_t^n \subseteq \mathfrak{D}_t^{n+1}$ and $\bigcup_n \mathfrak{D}_t^n = \mathfrak{D}_t$. Thus, by Lemma 2 we may approximate (PP) by a problem of the form

$$(\mathrm{PP}_n) \left\| \begin{array}{ll} \min_{x} x_0^\top S_0 \\[2mm] \text{s.t. } \mathbb{E}^{\mathbb{P}_0}\left[(-x_{t-1}^\top S_t + x_t^\top S_t + C_t) \cdot Z_t^{i,j} f_t^i \right] \leq 0 & \forall t < T; \forall i \leq n_1; \forall j \leq n_2^i \\[2mm] \mathbb{E}^{\mathbb{P}_0}\left[(-x_{T-1}^\top S_T + C_T) \cdot Z_T^{i,j} f_T^i \right] \leq 0 & \forall i \leq n_1; \forall j \leq n_2^i. \end{array} \right.$$

Rearranging its Lagrangian leads to the following representation of (PP_n):

$$\left\| \begin{array}{l} \inf_{x}\ \sup_{\lambda_0 \geq 0, \lambda_t^{i,j} \geq 0} \left\{ x_0^\top \left(\lambda_0 S_0 - \mathbb{E}^{\mathbb{P}_0}\left[S_1 W_1^n \right] \right) \right. \\[4mm] \qquad\qquad + \sum_{t=1}^{T-1} \mathbb{E}^{\mathbb{P}_0}\left[x_t^\top \left(S_t W_t^n - \mathbb{E}^{\mathbb{P}_0}\left[S_{t+1} W_{t+1}^n \middle| \mathcal{F}_t \right] \right) \right] \\[4mm] \qquad\qquad + \left. \sum_{t=1}^{T} \mathbb{E}^{\mathbb{P}_0}\left[C_t W_t^n \right] \right\}, \end{array} \right. \tag{14}$$

where

$$W_t^n := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2^i} \lambda_t^{i,j} Z_t^{i,j} f_t^j.$$

This is a finite-dimensional bilinear problem. Notice that (PP_n) is always feasible.[8] We may thus interchange the inf and the sup. Carrying out explicitly the minimiza-

---

[7] It would be sufficient to assume $\mathcal{Z}_{\mathcal{A}_t} \subseteq L_s$ and $f_t \in L_r$ such that $\frac{1}{r} + \frac{1}{s} = \frac{1}{q}$. However, for simplicity, we keep $\mathcal{Z}_{\mathcal{A}_t} \subseteq L_q$ and assume $f_t \in L_\infty$.

[8] This follows from the fact that a feasible solution $(x_0, \ldots, x_{T-1})$ of (PP_n) can easily be constructed in a deterministic way, starting with x_{T-1}.

tion in x, the unconstrained minimax problem (14) can be written as the constrained maximization problem

$$
\left\|
\begin{array}{l}
\sup\limits_{\lambda_t^{i,j} \geq 0} \sum\limits_{t=1}^{T} \mathbb{E}^{\mathbb{P}_0} \left[C_t W_t^n \right] \\[2ex]
\text{s.t. } S_t W_t^n = \mathbb{E}^{\mathbb{P}_0} \left[S_{t+1} W_{t+1}^n \middle| \mathcal{F}_t \right] \qquad \forall t = 1, \ldots, T \\[2ex]
\qquad W_t^n = \sum\limits_{i=1}^{n_1} \sum\limits_{j=1}^{n_2^i} \lambda_t^{i,j} Z_t^{i,j} f_t^j \qquad \forall t = 1, \ldots, T.
\end{array}
\right.
$$

Introducing a new probability measure $\mathbb{Q}$ defined by the Radon–Nikodým derivative $\frac{d\mathbb{Q}}{d\mathbb{P}_0} = W_T^n$, the problem can be rewritten in terms of $\mathbb{Q}$ in the form

$$
(\mathrm{DD}_n) \quad
\left\|
\begin{array}{l}
\sup\limits_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}} \left[\sum\limits_{t=1}^{T} C_t \right] \\[3ex]
\text{s.t. } \mathbb{E}^{\mathbb{Q}} \left[S_{t+1} \middle| \mathcal{F}_t \right] = S_t, \qquad \forall t = 0, \ldots, T-1 \\[2ex]
\qquad \left. \dfrac{d\mathbb{Q}}{d\mathbb{P}_0} \right|_{\mathcal{F}_t} \in \mathfrak{D}_t^n \,.
\end{array}
\right.
$$

It is left to show that there is no duality gap in the limit, as $n \to \infty$. Assume that the dual problem (DD) has an optimal value $\pi_a' \neq \pi_a$. By the primal constraints in (PP), for any dual feasible solution $\mathbb{Q}$ it holds

$$
\mathbb{E}^{\mathbb{Q}} \left[\sum_{t=1}^{T} C_t \right] \leq \mathbb{E}^{\mathbb{P}} \left[\sum_{t=1}^{T-1} (x_{t-1}^\top S_t - x_t^\top S_t) \cdot Z_t f_t + x_{T-1}^\top S_T \cdot Z_T f_T \right] = x_0^\top S_0.
$$

Thus, the optimal primal solution π_a is also greater than or equal to the optimal dual solution π_a'. Now assume $\pi_a' < \pi_a$. Then, since $\pi_a^n \uparrow \pi_a$ by Lemma 2, there must exist some n such that $\pi_a^n > \pi_a'$. Moreover, there exists some $\mathbb{Q}^n$, which is dual feasible and such that $\mathbb{E}^{\mathbb{Q}^n} \left[\sum_{t=1}^{T} C_t \right] = \pi_a^n$. This is a contradiction to π_a' being the limit of the monotonically increasing sequence of optimal values of the approximate dual problems of the form (DD_n). Hence, $\pi_a' = \pi_a$, i.e., it is shown that there is no duality gap in the limit.

Finally, considering the structure of $\mathfrak{D}_t$, the condition $\left. \frac{d\mathbb{Q}}{d\mathbb{P}_0} \right|_{\mathcal{F}_t} \in \mathfrak{D}_t$ means that it is of the form $Z_t f_t$, where there exists some $\mathbb{P} \in \mathcal{P}_\varepsilon$ such that $Z_t \in \mathcal{Z}_{\mathcal{A}_t^{\mathbb{P}}}$ and $\left. \frac{d\mathbb{P}}{d\mathbb{P}_0} \right|_{\mathcal{F}_t} = f_t$. This completes the derivation of the dual problem formulation (DD). $\square$

3.2 Nested distance balls as ambiguity sets: a large deviations result

In order to find appropriate nonparametric distances for probability models used in the framework of stochastic optimization, one has to observe that a minimal requirement is that it metricizes weak convergence and allows for convergence of empirical distributions. The Kantorovich–Wasserstein distance does metricize the weak topology

on the family of probability measures having a first moment. Its multistage generalization, the nested distance, measures the distance between stochastic processes on filtered probability spaces. The "Appendix" contains the definition and interpretation of both, the Kantorovich–Wasserstein distance and the nested distance.

Realistic probability models must be based on observed data. While for single- or vector-valued random variables with finite expectation the empirical distribution based on an i.i.d. sample converges in Kantorovich–Wasserstein distance to the underlying probability measure, the situation is more involved for stochastic processes. The simple empirical distribution for stochastic processes does not converge in nested distance (cf. Pflug and Pichler [32]), but a smoothed version involving density estimates does.

As we show here by merging the concepts of kernel estimations and transportation distances, one may get good estimates for confidence balls and ambiguity sets under some assumptions on regularity.

Let $\mathbb{P}$ be the distribution of the stochastic process $\xi = (\xi_1, \ldots, \xi_T)$ with values $\xi_t \in \mathbb{R}^m$. Notice that $\mathbb{P}$ is a distribution on $\mathbb{R}^\ell$ with $\ell = m \cdot T$. Let $\mathbb{P}^n$ be the probability measure of n independent samples from $\mathbb{P}$. If $\xi^{(j)} = (\xi_1^{(j)}, \ldots, \xi_T^{(j)})$, $j = 1, \ldots, n$ is such a sample, then the empirical distribution $\hat{\mathbb{P}}_n$ puts the weight $1/n$ on each of the paths $\xi^{(j)}$. For the construction of nested ambiguity balls, the empirical distribution has to be smoothed by convolution with a kernel function $k(x)$ for $x \in \mathbb{R}^\ell$. For a bandwidth $h > 0$ to be specified later, let $k_h(x) = \frac{1}{h^\ell} k(x/h)$. In what follows we will work with the kernel density estimate $\hat{f}_n = \hat{\mathbb{P}}_n * k_h$, where $*$ denotes convolution.

Assumption A4 1. The support of $\mathbb{P}$ is a set $D = D_1 \times \cdots \times D_T$, where D_i are compact sets in $\mathbb{R}^m$;
2. $\mathbb{P}$ has a Lebesgue density f, which is Lipschitz on D with constant L;
3. f is bounded from below and from above on D by $0 < \underline{c} \leq f(x) \leq \overline{c}$;
4. the kernel function k vanishes outside the unit ball and is Lipschitz with constant L;
5. the conditional probabilities $\mathbb{P}_t(A|x) = \mathbb{P}(\xi_t \in A | (\xi_1, \ldots, \xi_{t-1}) = x)$ satisfy

$$\mathsf{d}\left(\mathbb{P}_t\left(\cdot|x\right), \mathbb{P}_t\left(\cdot|y\right)\right) \leq \gamma_t \|x - y\|, \quad x, y \in D \tag{15}$$

for some $\gamma_t > 0$. Here, d denotes the Wasserstein distance for probabilities on $\mathbb{R}^m$.

Remark 2 The proof of Theorem 3 below relies on the lower bound $\underline{c}$ of the density. As the denominator of the conditional density $f(x|y) = f(x, y)/f(y)$ has to be estimated by density estimation as well, the bound ensures that the denominator does not vanish. In fact, the assumptions on the compact cube (point 1.) can be weakened to D being a compact set; the proof, however, is slightly more involved then. For the other technical assumptions (under point 5.) we may refer to Mirkov and Pflug [23].

Theorem 3 *(Large deviation for the nested distance) Under Assumption A4 there exists a constant $K > 0$ such that*

$$\mathbb{P}^n\left(\mathsf{dl}\left(\mathbb{P}, \hat{\mathbb{P}}_n * k_h\right) > \varepsilon\right) < \exp(-Kn\varepsilon^{2\ell+4}), \tag{16}$$

for n sufficiently large and appropriately chosen bandwidth h. Here, dl *denotes the nested distance.*

The proof of (16) is based on several steps presented as propositions below. To start with we recall two important results for density estimates $\hat{f}_n = \hat{\mathbb{P}}_n * k_h$ for densities f on $\mathbb{R}^\ell$.

Proposition 3 *Under the Lipschitz conditions for f and k given above, it holds that*

$$\mathbb{P}^n \left(\sup_{x \in D} \left| f(x) - \hat{f}_n(x) \right| > \varepsilon \right) \leq \mathbb{P}^n \left(\mathsf{d} \left(\hat{\mathbb{P}}_n, \mathbb{P} \right) > \left(\frac{\varepsilon}{2L} \right)^{\ell+2} \right). \tag{17}$$

if the bandwidth is chosen as $h = \varepsilon/(2L)$.

Proof See Bolley et al. [2, Prop. 3.1]. □

Proposition 4 *Let f and g be densities vanishing outside a compact set D and set $\mathbb{P}^f(A) = \int_A f(x)\mathrm{d}x$ resp. $\mathbb{P}^g(A) = \int_A g(x)\mathrm{d}x$. Then their Wasserstein distance d is bounded by*

$$\mathsf{d} \left(\mathbb{P}^f, \mathbb{P}^g \right) \leq 2\Delta\lambda(D) \, \|f - g\|_\infty. \tag{18}$$

Here Δ is the diameter of D and $\lambda(D)$ is the Lebesgue measure of D.

Proof Cf. [32, Prop. 4]. □

The next result extends the previous for conditional densities.

Proposition 5 *Let f and g be bivariate densities on compact sets $\bar{D}_1 \times \bar{D}_2$ bounded by $0 < \underline{c} \leq f, g \leq \bar{c} < \infty$ which are sufficiently close so that $\|f - g\|_{\bar{D}_1 \times \bar{D}_2} \leq \underline{c}\lambda(\bar{D}_1 \times \bar{D}_2)[2\Delta^\ell]^{-1}$. Then there is a universal constant κ_1, depending on the set $\bar{D} := \bar{D}_1 \times \bar{D}_2$ only, so that the conditional densities are close as well, i.e., they satisfy*

$$|f(x|y) - g(x|y)| \leq \kappa_1 \sup_{x' \in \bar{D}_1, y' \in \bar{D}_2} \left| f(x', y') - g(x', y') \right|$$

for all $x \in \bar{D}_1$ and $y \in \bar{D}_2$, i.e.,

$$\sup_{y \in \bar{D}_2} \|f(\cdot|y) - g(\cdot|y)\|_{\bar{D}_1} \leq \kappa_1 \|f - g\|_{\bar{D}_1 \times \bar{D}_2}. \tag{19}$$

Proof To abbreviate the notation set $\varepsilon := \sup_{x,y} |f(x, y) - g(x, y)|$ and note that $\varepsilon \leq \underline{c}\lambda(\bar{D})[2\Delta^\ell]^{-1}$. Consider the marginal density $f(y) := \int_{\bar{D}_1} f(x, y)\mathrm{d}x$ ($g(y) := \int_{\bar{D}_1} g(x, y)\mathrm{d}x$, resp.). It holds that

$$|f(y) - g(y)| \leq \int_{\bar{D}_1} |f(x, y) - g(x, y)| \, \mathrm{d}x \leq \int_{\bar{D}_1} \varepsilon \, \mathrm{d}x \leq \Delta^\ell \cdot \varepsilon.$$

Clearly $|f(y)| \geq \underline{c}\lambda(\bar{D}_1)$, where $\lambda(\bar{D}_1)$ is the Lebesgue measure of $\bar{D}_1$ and therefore

$$\left| \frac{f(y) - g(y)}{f(y)} \right| \leq \frac{\Delta^\ell}{\underline{c}\lambda(\bar{D}_1)} \cdot \varepsilon \leq \frac{1}{2}. \tag{20}$$

The elementary inequality $\frac{1}{1+x} \leq 1 + 2|x|$ is valid for $x \geq -1/2$. With (20) it follows that

$$
\begin{aligned}
g(x|y) - f(x|y) &= \frac{g(x, y)}{g(y)} - \frac{f(x, y)}{f(y)} = \frac{g(x, y)}{f(y)} \cdot \frac{1}{1 + \frac{g(y)-f(y)}{f(y)}} - \frac{f(x, y)}{f(y)} \\
&\leq \frac{g(x, y)}{f(y)} \left(1 + 2\frac{|g(y) - f(y)|}{f(y)} \right) - \frac{f(x, y)}{f(y)} \\
&= \frac{g(x, y) - f(x, y)}{f(y)} + 2\frac{g(x, y)}{f(y)} \frac{|g(y) - f(y)|}{f(y)} \\
&\leq \frac{\varepsilon}{\underline{c}\lambda(\bar{D}_1)} + 2\frac{\bar{c}}{\underline{c}\lambda(\bar{D}_1)} \frac{\Delta^\ell}{\underline{c}\lambda(\bar{D}_1)} \cdot \varepsilon \leq \kappa_1 \varepsilon
\end{aligned}
$$

with $\kappa_1 = \frac{1}{\underline{c}\lambda(\bar{D}_1)} + \frac{2\bar{c}\Delta^\ell}{(\underline{c}\lambda(\bar{D}_1))^2}$. The assertion of the proposition finally follows by exchanging the roles of the densities f and g. □

Theorem 4 *Given Assumption A4 there exists a constant κ_2 such that*

$$\mathbb{P}^n \left(\sup_{y \in \bar{D}_2} \mathsf{d} \left(\mathbb{P}^{f(\cdot|y)}, \mathbb{P}^{\hat{f}_n(\cdot|y)} \right) > \varepsilon \right) \leq \exp(-\kappa_2 n \varepsilon^{2\ell+4}) \tag{21}$$

for all $\varepsilon > 0$ and n sufficiently large.

Proof It follows from (18) and (19) that

$$\mathsf{d} \left(\mathbb{P}^{f(\cdot|y)}, \mathbb{P}^{\hat{f}_n(\cdot|y)} \right) \leq \kappa_3 \left\| f(\cdot|y) - \hat{f}_n(\cdot|y) \right\|_\infty \leq \kappa_3 \left\| f - \hat{f}_n \right\|_\infty$$

for $\kappa_3 = 2\Delta\lambda(D)\kappa_1$. Recall the large deviation result from [2, Th. 2.8], which is given by

$$\mathbb{P}^n(\mathsf{d}(\hat{\mathbb{P}}_n, \mathbb{P}) > \eta) \leq \exp(-n\kappa'\eta^2),$$

for some universal constant κ' depending on the Lipschitz constants of f and k only. With (17) it follows that

$$
\begin{aligned}
\mathbb{P} \left(\sup_{y \in \bar{D}_2} \mathsf{d} \left(\mathbb{P}^{f(\cdot|y)}, \mathbb{P}^{\hat{f}_n(\cdot|y)} \right) > \varepsilon \right) &\leq \mathbb{P} \left(\left\| f - \hat{f}_n \right\|_\infty > \frac{\varepsilon}{\kappa_3} \right) \\
&\leq \mathbb{P}^n \left(\mathsf{d} \left(\hat{\mathbb{P}}_n, \mathbb{P} \right) > \frac{\varepsilon^{\ell+2}}{(2L\kappa_3)^{\ell+2}} \right) \leq \exp \left\{ -\kappa'n \left(\frac{\varepsilon^{\ell+2}}{(2L\kappa_3)^{\ell+2}} \right)^2 \right\}.
\end{aligned}
$$

Setting $\kappa_2 := \kappa'(2L\kappa_3)^{-2\ell-4}$ in (21) reveals the result. ☐

Proof of Theorem 3 The previous theorem will be applied to the conditional densities of ξ_t given the past $\xi_1, \ldots, \xi_{t-1}$. Thus the sets $\bar{D}_i$ are interpreted as $\bar{D}_1 = D_t$ and $\bar{D}_2 = D_1 \times \cdots \times D_{t-1}$. For the probability measure $\mathbb{P}$ satisfying (15) and any other measure $\tilde{\mathbb{P}}$ satisfying $\mathsf{d}\left(\mathbb{P}_t\left(\cdot|x\right), \tilde{\mathbb{P}}_t\left(\cdot|x\right)\right) \le \varepsilon_t$ at stage t we have that

$$\mathsf{dl}\left(\mathbb{P}, \tilde{\mathbb{P}}\right) \le \sum_{t=1}^{T} \varepsilon_t \gamma_t \prod_{s=t+1}^{T}(1 + \gamma_s),$$

see [31, Sec. 4.2] or [23].

We employ the results elaborated above for $\tilde{\mathbb{P}} := \hat{\mathbb{P}}_n * k_h$. Then

$$\mathbb{P}^n\left(\mathsf{dl}\left(\mathbb{P}, \hat{\mathbb{P}}_n * k_h\right) > \varepsilon\right)$$

$$\le \mathbb{P}^n\left(\sum_{t=1}^{T} \mathsf{d}\left(\mathbb{P}_t\left(\cdot|x_t\right), \tilde{\mathbb{P}}_t\left(\cdot|x_t\right)\right) \gamma_t \prod_{s=t+1}^{T}(1 + \gamma_s) > \varepsilon\right)$$

$$= \sum_{t=1}^{T} \mathbb{P}^n\left(\mathsf{d}\left(\mathbb{P}_t\left(\cdot|x_t\right), \tilde{\mathbb{P}}_t\left(\cdot|x_t\right)\right) > \frac{\varepsilon}{T\gamma_t \prod_{s=t+1}^{T}(1 + \gamma_s)}\right).$$

We employ (21) to deduce that

$$\mathbb{P}^n\left(\mathsf{dl}\left(\mathbb{P}, \hat{\mathbb{P}}_n * k_h\right) > \varepsilon\right) \le \sum_{t=1}^{T} e^{-\kappa_2 n \varepsilon_t^{2\ell+4}}$$

with $\varepsilon_t := \varepsilon[T\gamma_t \prod_{s=t+1}^{T}(1 + \gamma_s)]^{-1}$.

The desired large deviation result follows for n sufficiently large for any $K < \min_{t \in \{1, \ldots, T\}} \kappa_2 \left[\left(T\gamma_t \prod_{s=t+1}^{T}(1 + \gamma_s)\right)^{2\ell+4}\right]^{-1}$. ☐

The smoothed model $\hat{\mathbb{P}}_n * k_h$ is not yet a tree, but by Theorem 6 of the "Appendix" one may find[9] a finite tree process $\bar{\mathbb{P}}_n$, which is arbitrarily close to it. Therefore, by eventually increasing the probability bound in (16) by another constant factor, it holds true also for $\bar{\mathbb{P}}_n$.

Remark 3 From a statistical perspective, the results contained in this section represent a strong motivation to use nested distance balls as ambiguity sets for general stochastic optimization problems on scenario trees constructed from observed data. In particular, the distributionally robust acceptable ask price allows the seller of a claim to invest in a trading strategy which gives an acceptable superhedge of the payments to be made under the *true* model with arbitrary high probability, given sufficient available data.

---

[9] See [31, Chap. 4] for methods to efficiently construct multistage models/scenario trees from data.

4 Illustrative examples

One may summarize the results of the previous sections in the following way: If the martingale measure is not unique ('incomplete market'), then typically there is a positive bid–ask spread in the (pointwise) replication model. This spread does also exist in the acceptability model. However, if the acceptability functional is the $\mathbb{AV@R}_\alpha$, then by changing α we can get the complete range between the replication model ($\alpha \to 0$) and the expectation model ($\alpha = 1$). At least in the latter case, but possibly even for some $\alpha < 1$, there is no bid–ask spread and thus a unique price. On the other hand, model ambiguity widens the bid–ask spread: The more models are considered, i.e., the larger the radius of the ambiguity set, the wider is the bid–ask spread. For illustrative purposes, let us look at the simplest form of examples which demonstrate these effects.

Example 1 Consider a three-stage ternary tree, where the paths are uniformly distributed and given by the columns of the matrix

$$\begin{bmatrix} 100 & 100 & 100 & 100 & 100 & 100 & 100 & 100 & 100 \\ 110 & 110 & 110 & 100 & 100 & 100 & 90 & 90 & 90 \\ 112 & 110 & 108 & 102 & 100 & 98 & 92 & 90 & 88 \end{bmatrix}.$$
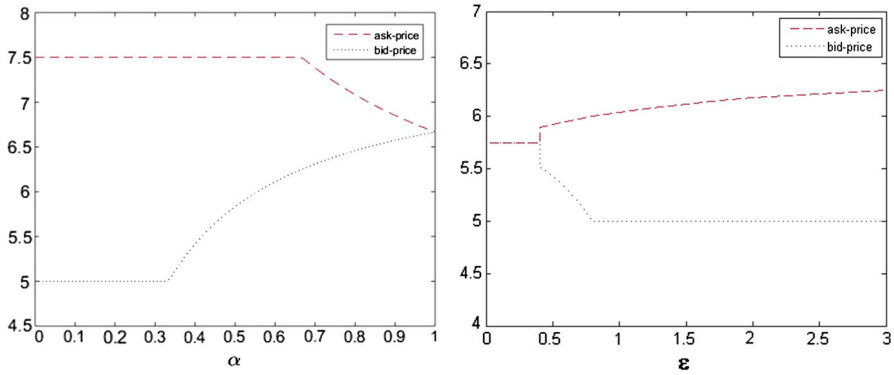
Since infinitely many equivalent martingale measures can be constructed on this tree, there is a considerable bid–ask spread for the pointwise replication model, which corresponds to the $\mathbb{AV@R}_\alpha$-acceptability pricing model with $\alpha = 0$. However, by increasing α for both contract sides, the bid–ask spread gets monotonically smaller. For $\alpha = 1$, there is no bid–ask spread, since all martingale measures coincide in their expectation and both buyer and seller only consider expectation in their valuation. Figure 1a visualizes this behavior for the price of a call option struck at 95%: the bid price increases with α, while the ask price decreases. For $\alpha = 1$ they coincide.

Computationally, $\mathbb{AV@R}$–acceptability pricing on scenario trees boils down to solving a linear program (LP). It is thus straightforward to implement and the problem scales with the complexity of LPs.

Example 2 In contrast, one may consider a three-stage binary tree model with uniformly distributed scenarios given by the columns of the matrix

$$\begin{bmatrix} 100 & 100 & 100 & 100 \\ 105 & 105 & 95 & 95 \\ 108 & 102 & 98 & 92 \end{bmatrix}.$$

This tree can carry only one single martingale measure. In such a model, the change of acceptability levels does not change the price, since also under weakened acceptability the price is determined by a martingale measure, namely the unique one (in case α is small enough such that it is feasible). However, in an ambiguity situation, a bid–ask spread may appear, since there are typically many martingale measures contained in ambiguity sets. We consider nested distance balls around the baseline tree, where we keep the uniform distribution of the scenarios for simplicity, but allow the

(a) Acceptability: The bid-ask spread tightens for increasing acceptability.

(b) Ambiguity: A bid-ask spread opens for increasing ambiguity.

Fig. 1 Distributionally robust acceptability pricing: the bid–ask spread as a function of the acceptability level α and the ambiguity radius ε

values of the process to change.[10] The result for a call option struck at 95% can be seen in Fig. 1b. While there is a unique price for small radii ε of the nested distance ball, an increasing bid–ask spread appears for larger values of ε.

5 Algorithmic solution

The nested distance between two given scenario trees can be obtained by solving an LP. However, the distributionally robust $\mathbb{AV@R}$–acceptability pricing problem w.r.t. nested distance balls as ambiguity sets results in a highly non-linear, in general non-convex problem. Therefore, we assume the tree structure to be given by the baseline model. In particular, it is assumed that different probability models within the ambiguity set differ only in terms of the transition probabilities; state values and the information structure are kept fixed.

Still, distributionally robust acceptability pricing is a semi-infinite non-convex problem. The only algorithmic approach available in the literature for similar problems is based on the idea of successive programming (cf. [31, Chap. 7.3.3]): an approximate solution is computed by starting with the baseline model only and alternately adding worst case models and finding optimal solutions. However, for typical instances of tree models this is computationally hard, as it involves the solution of a non-convex problem in each iteration step.

Hence, we tackle the dual formulation presented in Theorem 2. The structure of the nested distance enables an iterative approach. Algorithm 1 finds an approximate solution by solving a sequence of linear programs. Based on duality considerations and algorithmic exploitation of the specific stagewise transportation structure inherent

---

[10] This is a non-convex problem. The results in Fig. 1b are based on the standard nonlinear solver of a commercial software package (MATLAB 8.5 (R2015a), The MathWorks Inc., Natick, MA, 2015.), which finds (local) optima for our small instance of a problem.

to the nested distance, the algorithm approximates the solution of a semi-infinite non-convex problem by a sequence of LPs. The current state-of-the-art method, on the other hand, requires the solution of a non-convex program in each iteration step. Clearly, a sequential linear programming approach improves the performance considerably.[11] Moreover, our algorithm turned out to find feasible solutions in many cases where our implementation of a successive programming method fails to do so.

Let us extend the concept of the nested distance to subtrees, iteratively from the leaves to the root ('top-down'). For two scenario trees (here with identical filtration structures), define $\mathsf{dl}_T(i, j)$ as the distance of the paths leading to the leave nodes $i, j \in \mathcal{N}_T$. Moreover, define

$$\mathsf{dl}_t(k, l) := \sum_{i \in k+} \sum_{j \in l+} \pi(i, j | k, l) \, \mathsf{dl}_{t+1}(i, j),$$

for all nodes $k, l \in \mathcal{N}_t$, where $0 \leq t < T$. Then, the nested distance between the two trees is given by $\mathsf{dl}_0(1, 1)$. This stagewise backwards approach (cf. [31, Alg. 2.1]) is the basic idea of Algorithm 1. As we assume the tree structure to be fixed, Algorithm 1 iterates through the tree in the same top-down manner and searches for the optimal solution in each stage, while ensuring that the nested distance constraint remains satisfied. The variables are the conditional transition probabilities under $\mathbb{Q}$, i.e., $q_i := \mathbb{Q}[i | i-]$, as well as the transportation subplans $\pi(i, j | i-, j-)$, as defined in the "Appendix". We use the notation $n-$ for the immediate predecessor of some node n. As the measure $\mathbb{P}$ is in fact not needed explicitly since it is given by the transportation plan from $\hat{\mathbb{P}}$, condition (4.3) in Algorithm 1 serves to ensure that it is still well-defined implicitly (note that always some node $\tilde{k} \in \mathcal{N}_{t-1}$ needs to be fixed). Condition (1) ensures that $\mathbb{Q}$ is a martingale measure, $\mathbb{Q}$ represents conditional probabilities by condition (2), condition (3) corresponds to the constraint on the measure change ($d\mathbb{Q}/d\mathbb{P} \leq 1/\alpha$) resulting from the primal $\mathbb{AV@R}_\alpha$–acceptability conditions, and (4.1)–(4.3) represent the constraint that there must be one $\mathbb{P}$ contained in the nested distance ball such that condition (3) holds.

The algorithm optimizes the variables stagewise top-down. The optimal solution at stage $t + 1$ depends on the values of the variables for all stages up to stage t, which result from the previous iteration step. Therefore, the algorithm iterates as long as there is further improvement possible at some stage, given updated variable values for the earlier stages of the tree. Otherwise, it terminates and the optimal solution of our approximate problem is found.

Algorithm 1 Acceptability pricing under model ambiguity.

Start with some feasible model $\mathbb{P}$ in the nested distance ball around $\hat{\mathbb{P}}$. Initialize π_{old} by assigning the optimal transportation plan between $\mathbb{P}$ and $\hat{\mathbb{P}}$ and initialize 'oldprice'.

1: **Iteration**
2:     [newprice, π_{new}] ← GETPRICE(π_{old})
3:     **if** (oldprice == newprice) **then**
4:         **return** oldprice
5:     **else**

---

[11] For our implementations, the speed-up factor for a test problem was on average about 100. However, this may depend heavily on the implementation and the problem.

6: oldprice $\leftarrow$ newprice, $\pi_{\text{old}} \leftarrow \pi_{\text{new}}$
7: **Iterate**
8: **end if**
9: **EndIteration**

10: **function** GETPRICE($\tilde{\pi}$)
11: **for** t **from** T **to** 1 **do solve**

$$\max_{\{q_i, \, \pi(i,j|k,l) \, : \, i,j \in \mathcal{N}_t\}} \mathbb{E}^{\mathbb{Q}}\left[\sum_{\tau=t}^{T} C_\tau \middle| \mathcal{F}_{t-1}\right]$$

$$\text{s.t.}$$

$$(1) \quad \sum_{i \in k+} q_i \cdot x_i = x_k \qquad \forall k \in \mathcal{N}_{t-1}$$

$$(2) \quad \sum_{i \in k+} q_i = 1 \qquad \forall k \in \mathcal{N}_{t-1}$$

$$(3) \quad -\sum_{i \in k+} \pi(i,j|k,l) + \alpha \cdot q_j \le 0 \qquad \forall j \in \mathcal{N}_t$$

$$(4.1) \quad \sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t} \pi(i,j|i-,j-) \cdot \tilde{\pi}(i-,j-) \cdot \mathsf{dl}_t(i,j) \le \varepsilon$$

$$(4.2) \quad \sum_{j \in l+} \pi(i,j|k,l) = \hat{\mathbb{P}}[i|k] \qquad \forall l \in \mathcal{N}_{t-1}, \forall i \in \mathcal{N}_t$$

$$(4.3) \quad \sum_{i \in k+} \pi(i,j|k,l) = \sum_{i \in \tilde{k}+} \pi(i,j|\tilde{k},l) \quad \forall k \in \mathcal{N}_{t-1}, \forall j \in \mathcal{N}_t$$

$$(5) \quad q_i, \pi(i,j|i-,j-) \in [0,1] \qquad \forall i,j \in \mathcal{N}_t$$

12: **end for**
13: price $\leftarrow \mathbb{E}^{\mathbb{Q}}[\sum_{t=1}^{T} C_t]$, construct transportation plan $\pi(\cdot,\cdot)$ from subplans $\pi(\cdot,\cdot|\cdot,\cdot)$
14: **return** [price, π]
15: **end function**

Example 3 Consider the price of a plain vanilla call option struck at 95, in the Black–Scholes model with parameters $S_0 = 100, r = 0.01, \sigma = 0.2, T = 1$. Applying optimal quantization techniques (see, e.g., [31, Chap. 4] for an overview) to discretize the lognormal distribution, we construct a scenario tree with 500 nodes. While there exists a unique martingale measure (and thus a unique option price) in the Black–Scholes model, the discrete approximation allows for several martingale measures (and thus a positive bid–ask spread). Figure 2 visualizes the bid–ask spread as a function of the $\mathbb{AV}@R$–acceptability level α and the radius ε of the nested distance ball used as model ambiguity set. For $\alpha \to 1$ and $\varepsilon = 0$, the spread closes and the resulting price approximates the true Black–Scholes price up to 4 digits. For illustrative purposes, the spread between the bid and the ask price surface is shown from two perspectives.

6 Conclusion

In this paper we extended the usual methods for contingent claim pricing into two directions. First, we replaced the replication constraint by a more realistic acceptability constraint. By doing so, the claim price does explicitly depend on the stochastic model for the price dynamics of the underlying (and not just on its null sets). If the model is based on observed data, then the calculation of the claim price can be seen as a statistical
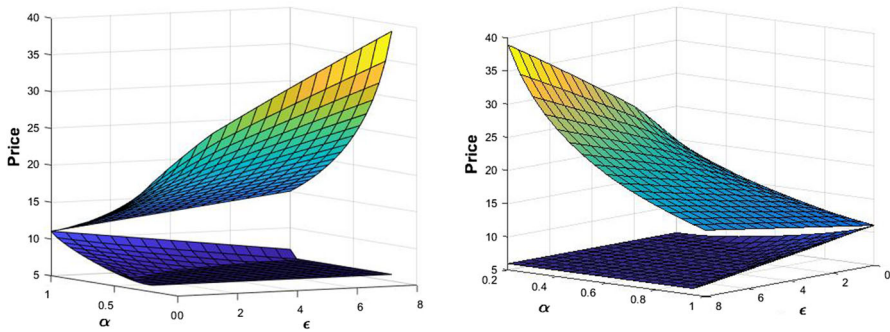
Fig. 2 The bid–ask spread as a function of acceptability and ambiguity

estimate. Therefore, as a second extension, we introduced model ambiguity into the acceptability pricing framework and we derived the dual problem formulations in the extended setting. Moreover, we used the nested distance for stochastic processes to define a confidence set for the underlying price model. In this way, we link acceptability prices of a claim to the quality of observed data. In particular, the size of the confidence region decreases with the sample size, i.e., the number of observed independent paths of the stochastic process of the underlying. For a given sample of observations, the ambiguity radius indicates how much the baseline ask/bid price should be corrected to safeguard the seller/buyer of a claim against the inherent statistical model risk, as Sect. 5 illustrates.

Appendix

Distances for random variables and stochastic processes. Recall the definition of the Kantorovich–Wasserstein distance $\mathsf{d}(P, \tilde{P})$ for two (Borel) random distributions P and $\tilde{P}$ on $\mathbb{R}^m$:

$$
\begin{aligned}
\mathsf{d}(P, \tilde{P}) := \inf_{\pi} &\iint \|\omega - \tilde{\omega}\| \, \pi(d\omega, d\tilde{\omega}) \\
\text{s.t. } &\pi\left(A \times \mathbb{R}^m\right) = P(A) \\
&\pi\left(\mathbb{R}^m \times B\right) = \tilde{P}(B).
\end{aligned}
$$

Here, π runs over all Borel measures on $\mathbb{R}^m \times \mathbb{R}^m$ with given marginals P resp. $\tilde{P}$. These measures are called *transportation plans*. If ξ and $\tilde{\xi}$ are $\mathbb{R}^m$-valued random variables, then their distance is defined as the distance of the corresponding image measures P^{ξ} resp. $P^{\tilde{\xi}}$.

Pflug and Pichler [29,30] introduced the notion of the nested distance as a generalization of the Kantorovich–Wasserstein distance for $\mathbb{R}^m$-valued stochastic processes $\xi = (\xi_1, \ldots, \xi_T)$ and its image measures $\mathbb{P}$ on $\mathbb{R}^{mT}$. Let $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_T)$ be the filtration composed of the sigma-algebras $\mathcal{F}_t$ generated by the component projections $(\xi_1, \ldots, \xi_T) \mapsto (\xi_1, \ldots, \xi_t)$. Moreover, let for $\xi = (\xi_1, \ldots, \xi_T) \in \mathbb{R}^{mT}$ the distance be defined as $\|\xi - \tilde{\xi}\| := \sum_{t=1}^{T} \|\xi_t - \tilde{\xi}_t\|$.

Definition 3 The nested distance dl for distributions $\mathbb{P}$ and $\tilde{\mathbb{P}}$ is defined as

$$
\left\|
\begin{aligned}
\mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}) &:= \inf_{\pi} \iint \|\xi(\omega) - \tilde{\xi}(\tilde{\omega})\| \, \pi(d\omega, d\tilde{\omega}) \\
\text{s.t. } \pi &\left(A \times \mathbb{R}^m \,\middle|\, \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t \right) = \mathbb{P}\left[A \,\middle|\, \mathcal{F}_t \right] \qquad A \in \mathcal{F}_T; \ t = 1, \ldots, T \\
\pi &\left(\mathbb{R}^m \times B \,\middle|\, \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t \right) = \tilde{\mathbb{P}}\left[B \,\middle|\, \tilde{\mathcal{F}}_t \right] \qquad B \in \tilde{\mathcal{F}}_T; \ t = 1, \ldots, T.
\end{aligned}
\right.
$$

To interpret this definition, the nested distance between two multistage probability distributions is obtained by minimizing over all transportation plans π, which are compatible with the filtration structures. For a single period (i.e., $T = 1$), the nested distance coincides with the Kantorovich–Wasserstein distance. The following basic theorem for stability of multistage stochastic optimization problems was proved by Pflug and Pichler [30, Th. 6.1].

Theorem 5 *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be nested distributions with filtrations $\mathcal{F}$ and $\tilde{\mathcal{F}}$, respectively. Consider the multistage stochastic optimization problem*

$$
v(\mathbb{P}) := \inf \left\{ \mathbb{E}^{\mathbb{P}}[Q(\xi, x)] \colon x \in \mathbb{X}, \, x \lhd \mathcal{F} \right\},
$$

where Q is convex in the decisions $x = (x_1, \ldots, x_T)$ for any ξ fixed, and Lipschitz with constant L in the scenario process $\xi = (\xi_1, \ldots, \xi_T)$ for any x fixed. The set $\mathbb{X}$ is assumed to be convex and the constraint $x \lhd \mathcal{F}$ means that the decisions can be random variables, but must be adapted to the filtration $\mathcal{F}$, i.e., must be nonanticipative. Then the objective values $v(\mathbb{P})$ and $v(\tilde{\mathbb{P}})$ satisfy

$$
\left| v(\mathbb{P}) - v(\tilde{\mathbb{P}}) \right| \le L \cdot \mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}).
$$

Finite scenario trees are much easier to work with than general stochastic processes. For finite trees, where every node m has a unique predecessor, we write $m+$ for the set of its immediate successors. Denote by $\mathcal{N}_t$ the set of all nodes at stage t of the tree model $\mathbb{P}$. For a node $i \in m+$ let $\mathbb{P}[i|m]$ be the conditional transition probability from m to i.

Definition 4 The nested distance for scenario trees $\mathbb{P}$ and $\tilde{\mathbb{P}}$ is defined as

$$
\left\|
\begin{aligned}
\mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}) := &\min_{\pi \geq 0} \sum_i \sum_j \pi_{i,j} \cdot D_{i,j} \\
&\text{s.t.} \sum_{j \in l+} \pi(i, j | k, l) = \mathbb{P}[i|k] \quad \forall i \in k+; \forall (k, l) \in (\mathcal{N}_t \times \tilde{\mathcal{N}}_t); 1 \leq t < T \\
&\qquad \sum_{i \in k+} \pi(i, j | k, l) = \tilde{\mathbb{P}}[j|l] \quad \forall j \in l+; \forall (k, l) \in (\mathcal{N}_t \times \tilde{\mathcal{N}}_t); 1 \leq t < T \\
&\qquad \pi_{i,j} \geq 0 \text{ and } \sum_i \sum_j \pi_{i,j} = 1.
\end{aligned}
\right\| \quad (22)
$$

The matrix π of transportation plans and the matrix D carrying the pairwise distances of the paths are defined on $\mathcal{N}_T \times \tilde{\mathcal{N}}_T$. The conditional joint probabilities $\pi(i, j | k, l)$ in (22) are given by $\pi(i, j | k, l) = \pi_{i,j} \cdot [\sum_{i' \in k+} \sum_{j' \in l+} \pi_{i',j'}]^{-1}$.

Approximation of random processes by finite trees The subsequent result follows from [31, Prop. 4.26].

Theorem 6 *If the stochastic process* $\xi = (\xi_1, \ldots, \xi_T)$ *satisfies the Lipschitz condition given in Assumption A4.5 in Sect.* 3.2, *then for every* $\varepsilon > 0$ *there is a stochastic process with distribution* $\tilde{\mathbb{P}}$, *which is defined on a finite tree and which satisfies*

$$
\mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}) \leq \varepsilon,
$$

where $\mathbb{P}$ *is the distribution of* ξ *on the filtered space* $(\Omega, \mathcal{F})$.

References

1. Analui, B., Pflug, G.Ch.: On distributionally robust multiperiod stochastic optimization. Comput. Manag. Sci. **11**(3), 197–220 (2014)
2. Bolley, F., Guillin, A., Villani, C.: Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Relat. Fields **137**(3–4), 541–593 (2007)
3. Carr, P.P., Geman, H., Madan, D.B.: Pricing and hedging in incomplete markets. J. Financ. Econ. **62**(1), 131–167 (2001)
4. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. Chapman & Hall/CRC, Boca Raton (2004)
5. Dahl, K.R.: A convex duality approach for pricing contingent claims under partial information and short selling constraints. Stoch. Anal. Appl. **35**(2), 317–333 (2017)
6. Delbaen, F., Schachermayer, W.: A general version of the fundamental theorem of asset pricing. Math. Ann. **300**(3), 463–520 (1994)
7. Duan, C., Fang, W., Jiang, L., Yao, L., Liu, J.: Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric. IEEE Trans. Power Syst. **PP**, 1 (2018)
8. Esfahani, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Math. Program. **171**(1–2), 115–166 (2018)
9. Föllmer, H., Leukert, P.: Quantile hedging. Finance Stoch. **3**(3), 251–273 (1999)
10. Föllmer, H., Leukert, P.: Efficient hedging: cost versus shortfall risk. Finance Stoch. **4**(2), 117–146 (2000)
11. Gao, R., Kleywegt, A.J.: Distributionally robust stochastic optimization with Wasserstein distance (2016)
12. Hanasusanto, G., Kuhn, D.: Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. Oper. Res. **66**(3), 849–869 (2018)

13. Harrison, J.M., Kreps, D.: Martingales and arbitrage in multiperiod securities markets. J. Econ. Theory **20**(3), 381–408 (1979)
14. Harrison, J.M., Pliska, S.R.: Martingales and stochastic integrals in the theory of continuous trading. Stoch. Process. Appl. **11**(3), 215–260 (1981)
15. Harrison, J.M., Pliska, S.R.: A stochastic calculus model of continuous trading: complete markets. Stoch. Process. Appl. **15**(3), 313–316 (1983)
16. Jouini, E., Schachermayer, W., Touzi, N.: Law Invariant Risk Measures have the Fatou Property, pp. 49–71. Springer, Tokyo (2006)
17. Kallio, M., Ziemba, W.T.: Using Tucker's theorem of the alternative to simplify, review and expand discrete arbitrage theory. J. Bank. Financ. **31**(8), 2281–2302 (2007)
18. King, A., Korf, L.: Martingale pricing measures in incomplete markets via stochastic programming duality in the dual of L^∞ (2002)
19. King, A.J.: Duality and martingales: a stochastic programming perspective on contingent claims. Math. Program. **91**(3), 543–562 (2002)
20. King, A.J., Koivu, M., Pennanen, T.: Calibrated option bounds. Int. J. Theor. Appl. Finance (IJTAF) **08**(02), 141–159 (2005)
21. King, A.J., Streltchenko, O., Yesha, Y.: Private Valuation of Contingent Claims in a Discrete Time/State Model, Chap. 27, pp. 691–710. Springer, Boston (2010)
22. Kreps, D.M.: Arbitrage and equilibrium in economies with infinitely many commodities. J. Math. Econ. **8**(1), 15–35 (1981)
23. Mirkov, R., Pflug, G.Ch.: Tree approximations of dynamic stochastic programs. SIAM J. Optim. **18**(3), 1082–1105 (2007)
24. Nakano, Y.: Efficient hedging with coherent risk measure. J. Math. Anal. Appl. **293**(1), 345–354 (2004)
25. Nguyen, V.A., Kuhn, D., Esfahani, P.M.: Distributionally robust inverse covariance estimation: the Wasserstein shrinkage estimator. Available from Optimization Online (2018)
26. Pennanen, T.: Convex duality in stochastic programming and mathematical finance. Math. Oper. Res. **36**, 340–362 (2011)
27. Pennanen, T.: Optimal investment and contingent claim valuation in illiquid markets. Finance Stoch. **18**(4), 733–754 (2014)
28. Pennanen, T., King, A.J.: Arbitrage pricing of American contingent claims in incomplete markets: a convex optimization approach. Stoch. Program. E-Print Ser. (2004)
29. Pflug, G.Ch.: Version-independence and nested distributions in multistage stochastic optimization. SIAM J. Optim. **20**(3), 1406–1420 (2009)
30. Pflug, G.Ch., Pichler, A.: A distance for multistage stochastic optimization models. SIAM J. Optim. **22**(1), 1–23 (2012)
31. Pflug, G.Ch., Pichler, A.: Multistage Stochastic Optimization, Springer Series in Operations Research and Financial Engineering, 1st edn. Springer, Berlin (2014)
32. Pflug, G.Ch., Pichler, A.: From empirical observations to tree models for stochastic optimization: convergence properties. SIAM J. Optim. **26**(3), 1715–1740 (2016)
33. Pflug, G.Ch., Römisch, W.: Modeling, Measuring and Managing Risk. World Scientific, Singapore (2007)
34. Pflug, G.Ch., Wozabal, D.: Ambiguity in portfolio selection. Quant. Finance **7**(4), 435–442 (2007)
35. Rockafellar, R.T.: Conjugate Duality and Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA (1974)
36. Rockafellar, R.T., Wets, R.J.-B.: Nonanticipativity and L^1-martingales in stochastic optimization problems. Math. Program. Study **6**, 170–187 (1976)
37. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: basic duality. Pac. J. Math. **62**(1), 173–195 (1976)
38. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: relatively complete recourse and induced feasibility. SIAM J. Control Optim. **14**(3), 574–589 (1976)
39. Rockafellar, R.T., Wets, R.J.-B.: Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. Pac. J. Math. **62**(2), 507–522 (1976)
40. Rockafellar, R.T., Wets, R.J.-B.: Measures as Lagrange multipliers in multistage stochastic programming. J. Math. Anal. Appl. **60**(2), 301–313 (1977)
41. Rockafellar, R.T., Wets, R.J.-B.: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. SIAM J. Control Optim. **16**, 16–36 (1978)
42. Rudloff, B.: Convex hedging in incomplete markets. Appl. Math. Finance **14**(5), 437–452 (2007)

43. Sarykalin, S., Serraino, G., Uryasev, S.: Value-at-risk vs. conditional value-at-risk in risk management and optimization. In: Tutorials in Operations Research, INFORMS, pp. 270–294 (2008). ISBN 978-1-877640-23-0
44. Scarf, H.: A Min–Max Solution of an Inventory Problem. Rand Corporation, Santa Monica (1957)
45. Van Parys, B.P., Esfahani, P.M., Kuhn, D.: From data to decisions: distributionally robust optimization is optimal. Available from Optimization Online (2017)
46. Zhao, C., Guan, Y.: Data-driven risk-averse stochastic optimization with Wasserstein metric. Oper. Res. Lett. **46**(2), 262–267 (2018)